EL COLEGIO DE MÉXICO, A.C.
CENTRO DE ESTUDIOS ECONÓMICOS

A DYNAMIC APPROACH TO MIGRATION AND VOTING PROBLEMS

TESIS PRESENTADA POR:

FERNANDA HERRERA LÓPEZ

PARA OPTAR POR EL GRADO DE

DOCTORA EN ECONOMÍA

PROMOCIÓN 2015-2018

DIRECTOR DE TESIS

DAVID RENÉ MICHEL CANTALA

CIUDAD DE MÉXICO ENERO DE 2019

EL COLEGIO DE MÉXICO

CENTRO DE ESTUDIOS ECONÓMICOS

CONSTANCIA DE APROBACIÓN

Doctorante:	Fernanda Herrera López
Tesis:	*A dynamic approach to migration and voting problems*
Director de Tesis:	Dr. David René Michel Cantala

Aprobada por el Jurado Examinador:

Dr. David René Michel Cantala	Presidente	________________________
Dr. Gilles Alexandre Paul Serra Baños	Primer Vocal	________________________
Dr. Edwin van Gameren	Vocal Secretario	________________________
Dra. Diana Terrazas Santamaría	Lector	________________________
Dr. Eneas Caldiño García	Suplente	________________________

Ciudad de México, 31 de enero de 2019

A El Colegio de México

Table of Contents

Labor migrant networks: Growth, saturation, and deflection to new labor markets.

Abstract

We present a stochastic process that models the time evolution of a migrant network. The process simultaneously captures two conflicting effects inherent to the expansion of the network: the increase in the job search efficiency and the reduction of vacancies at destination. In our model, potential migrants decide whether to use a network to search for jobs in a given destination or migrate to a different location (and start a network there). Our purpose is to identify the network size that causes the dispersion of migration. Findings show that when the network members have incomplete information on job availability, the deflection may occur before the destination runs out of vacancies. It is also shown that in some cases, the gains of network-driven migration are increased when creating networks before the critical size is reached.

1 Introduction

The theory of network migration highlights the role that social networks play in the migratory decision-making. The crucial assumption of this theory is that networks have the ability to reduce the costs and risks of migration through the transmission of information and other resources between its members (Massey *et al.*, 1990). The reduction in the migrating costs increases the expected value of migrating and this encourages aspiring migrants in the origin to out-migrate.[1] In turn, the newcomers add to the network's resource stock and reduce the migrating costs of future immigrants. Needles to say, this prompts additional immigration. This is the basis of the cumulative causation of migration,[2] the main premise is that "each act of migration alters the social context within which subsequent migration decisions are made, typically in ways that make additional movement more likely" (Massey, *et al.* 1998: 45-46).

A remarkable feature of this cumulative process is that it is able to change the composition of the immigration flow. Although migration initially draws individuals from a specific part of the skill (or income) distribution (Borjas, 2014), the use of social networks allows a variety of people from the sending community to migrate (Durand and Massey, 2003). For instance, McKenzie and Rapoport (2010) show that US destinations with low migrant networks and high migration costs attract Mexican migrants with higher education, while communities with high networks and low costs attract less educated individuals. Consequently, Massey (1990) inferred that the expansion of networks may induce a self-perpetuating migration process independent of the causes that originated migration in the first place.

In the late 1990's, scholars began questioning the limits to the theories of migration network and cumulative causation of migration. Massey, *et al.* (1998) advanced that at some point in the expansion of networks, the marginal decrease in the migrating costs is not enough to encourage individuals to out-migrate their source communities, while Fussell and Massey (2004) found that the effectiveness of the cumulative causation depends on the quality of the ties that link the network members. Other authors focused on the effect of the immigration

[1]Empirical evidence includes Taylor (1984), Massey *et al.* (1987), Massey and Espinosa (1997), and McKenzie and Rapoport (2007).

[2]This construction is adapted from Myrdal's theory of circular cumulative causation (1957).

influx on the destination's wages and resources and suggested that the reduction of economic opportunities available to immigrants counteracts the effects of cumulative causation. Heer (2002) showed that a high percentage of Hispanics in US destinations reduced the earnings of Hispanics immigrants and lowered subsequent immigration influxes. He observed that in this stage, the role of migrant networks is to deter immigration to unfavorable destinations and to redirect potential migrants to other locations.[3] Light (2006) criticized the fact that migration network theory ignores the reduction of resources that result from network expansion. More specifically, he argued that mature networks expose available jobs for immigrants, but if there is no increase in the supply of jobs, they also aggravate job shortages in destinations: "the contradiction is a fundamental one: migration networks enhace search efficiency but diminish resource availability" (Light, 2006: 82). A common factor in these observations is that when the benefits of network-driven migration decrease, immigration to saturated destinations is ceased or reduced.

It is of interest to know up to what point migrant networks encourage immigration to a destination, and as a corollary, what network size deflects migration to new destinations. To do so, we present a model in which potential migrants have the option of using a migrant network to search for jobs in a given destination or migrating to a different location without the use of a network. Every network size has a value of continued search and the critical value of a network is given by the size for which the value of search is lower than the value of migrating by oneself. Although the impact of social networks on search behavior and its outcome has been addressed before,[4] a novel factor in this study is that we explicitly introduce the effect of the reduction of vacancies in the network's search intensity. As a result, the evolution of the migrant network reflects the trade-off between job search efficiency

[3]Monras (2015) showed that the low-skilled immigrant supply shock brough about by the 1994 Mexican Peso Crisis decreased low-skilled wages in high immigration US states and prompted labor relocation across the country.

[4]See for instance Caliendo, *et al.* (2010), and Calvó-Armengol and Jackson (2007).

and resource reduction. Our main finding is that potential migrants avoid migrating to the targeted destination before it runs out of vacant jobs for newcomers. We also show that in some cases it may be more efficient for the whole immigrant population to open new networks before our model predicts they should.

This paper adds to the literature in two ways. First, we make a tractable definition of the economic saturation of networks. The new definition enables the computation of the critical size of a network given some economic variables from the source and host destinations. As a consequence, the model allows to predict saturation before it occurs, which is something previous studies have failed to do. Second, the stochastic nature of our model makes possible to draw a time dimension to the study of the expansion of social networks. We are thus able to compute the distribution of the saturation time of a network, a factor that to our knowledge of the literature has been neglected so far.

2 Model

The aim of this model is to analyze the impact of social networks on the migrating decisions. In particular, we are interested in finding the size of networks for which the benefits of network-driven migration are lower than the value of migrating by oneself. This critical size represents the *economic saturation point* of migrant networks. Given that migration networks are believed to be more useful the more uncertain the destination is (Taylor, 1986), we model international migration decisions. However, the setting of the model is general enough to apply it to internal migration.

The model relies on the following assumptions. There are two countries. The destination country has several locations with identical segmented labor markets, each with an exogenous demand for immigrant labor. The number of jobs for immigrants in each location is a fixed number N. We assume that potential migrants are identical. An agent in the source country has the option of staying in the origin, using a network to search for a job in a given destination, or migrating on her own to a different location and starting a network there. Her decision is based on the net discounted value of income across locations.

The key condition that we impose is that the network members do not know where vacancies are. This implies that they must engage in search to find jobs for potential immigrants. If an agent is willing to search for a job through a network, she pays the search cost incurred by the network members and gets the present value of income from the search. Two outcomes may arise. Either the network finds a vacancy, in which case the agent takes it and joins the network, or no available job is found, and the agent remains in the origin and decides whether to stay, use a network or migrate to a different location.

We are interested in modeling the impact of an increase of the network size on the value of search through a network. We will restrict our attention to the effect of the expansion on the availability of jobs at destination rather than on the wages of immigrants. This is because the host wages may be affected by several factors other than the volume of labor supply, so the effect of the expansion might be overshadowed by other variables. We assume that once an individual has been incorporated to a network, she does not leave it and keeps her job forever. This implies that the employment status of the network members is not affected by the arrival of new immigrants, and consequently they have no incentives to alter their search behavior as the network expands. We also assume that the only way to migrate to a location with a network is by means of the network. This assumption allows us to isolate the effects of network growth on the destination's resources. Note that it also allows us to use the terms immigrant and network member interchangeably.

The basic idea is that every immigrant that is added to the network increases the job search efficiency and at the same time reduces the number of available jobs in the destination.[5] This means that an increase of the network size reduces the average time in which jobs are sampled and simultaneously prolongs the expected time in which available jobs are selected. The overall time in which the network finds an available job for an immigrant (and

[5]For simplicity we assume that networks do not create jobs at destination.

adds a member) is given by the interaction of these conflicting effects. The advantage of modeling the transitions between network sizes by means of a stochastic process lies in the fact that the value of search may be easily expressed in a Bellman equation.

In Section 2.1 we set up notation and terminology for the time-evolution process of the network. It is worth pointing out that this process provides an intrinsic characterization of the network search intensities, but makes no appeal to the economics involved in the migrating decision. The principal significance of the network process is that it allows one to describe the transitions between the different network sizes. Section 3 indicates how the network intensities may be used to compute the value of search for every network size, it also presents the values of staying in the home country and starting a network in a host destination. In Section 4 we make a formal definition of the economic saturation of networks in terms of these expected values and provide an example. We also address the concept of network saturation time and derive its density and distribution functions. Section 5 discusses whether reaching the saturation point of networks is optimal for the immigrant population. We conclude in Section 6.

2.1 Network process

We call *network* a non-empty set of immigrants in a given destination, from this definition it is understood that a network may be composed of one immigrant. Let $\{X(t) : t \geq 0\}$ be the counting process of immigrants in the network. In other words, $X(t)$ gives the size of the network at time t. Under the condition that immigrants are invited to join the network only if there are vacant jobs for them, the maximal size of the network is given by the total number of jobs for immigrants N. Therefore, the state space of $\{X(t)\}_{t\geq 0}$ is the subset of natural numbers $S = \{1, 2, \ldots, N\}$.

Note that since immigrants are not allowed to leave the network, the number of network members is a non-decreasing function of time. We prove that $\{X(t)\}_{t\geq 0}$ is a *birth process*[6] with rates $q_1, \ldots, q_N$, where q_n gives the net effect of search efficiency improvement and reduction of vacancies when the network size is $n \in S$. That is, we show that the network size increases by one, and that transition from size n to $n+1$ occurs in a random exponential time of parameter q_n.

In what follows we show how the network rates $(q_i)_{i\in S}$ are obtained. Since we are interested in describing all the state transitions, we begin our study with a network of size one and work our way up to the maximum size N. We emphasize that our interest is to study how the addition of an immigrant to the network affects the time in which the next immigrant arrives, so for convenience we assume that all the potential immigrants are willing to use and join the network. In Section 3 we show how to dispense with the assumption.

Let us suppose that a potential immigrant prefers to migrate to a host location with no network, and thereby start one. This pioneer migrant secures one of the N available jobs with probability one. We measure time from the moment of her arrival, so we have $X(0) = 1$. The arrival of the next immigrant is given by the time in which the pioneer finds a vacancy. This variable is affected by the search intensity and the number of available jobs left. We study each factor separately, and then show how their interaction gives rise to the network's birth rates.

2.1.1 Job search efficiency

We assume that all the network members search for jobs for potential immigrants. Each member has a Poisson process of search and the search processes are independent amongst each other. The realizations of a search process are *observations* of jobs. An observation consists on the random selection of a job other than the own. Here, the underlying assumption is that immigrants know that their job is taken (by them), so they do not sample their own job when looking for a vacancy. Thus, the observations have a uniform distribution on

[6]See Norris, 1998.

the rest of the $N-1$ jobs.

An observed job can be available or taken. If it is available, the network invites an aspiring immigrant to take the job and join the network. In turn, this new member will search for jobs for other potential immigrants. If the observed job is already filled, the network makes no invitation and by the assumption that immigration to a destination with a network is possible only through it, the number of network members remains the same.

Immigrants have different search intensities that depend on their order of arrival.[7] The intensity of the nth immigrant is given by $\lambda\delta^{n-1}$, where λ is a positive constant, $0 < \delta < 1$ is a parameter of reduction and n in $\{1, \ldots, N\}$ denotes the order of arrival to the network. So, the search intensities are: λ for the pioneer immigrant, $\lambda\delta$ for the second, $\lambda\delta^2$ for the third, and so forth. Note that the intensities do not depend on time, so the search processes are homogeneous. This means that a network member will search with the same intensity throughout time.

By the definition of Poisson process, the times between job observations -also known as interarrival times- are exponentially distributed. Accordingly, high search intensities correspond to short expected interarrival times, and viceversa. On account of the above remark, the first immigrants take on average less time to observe jobs than new members. This is in line with Todaro's (1980) propositon that longer-term arrived migrants have more contacts and better information systems than the newly arrived. The condition can be relaxed since we can make δ arbitrarily close to one.

The network search process is composed of all the job observations made by the network members. To study the general case, let $X(t) = n$ for some n in S and $t \geq 0$. This means there are n independent search processes with respective parameters $\lambda, \lambda\delta, \ldots, \lambda\delta^{n-1}$. By the Superposition theorem,[8] the search process of the network is also a Poisson process of

[7]The mathematical convenience of making this assumption is more extensively discussed in Appendix B.
[8]See Kingman, 1993.

intensity

$$\lambda_n = \sum_{i=1}^{n} \lambda \delta^{i-1} = \lambda \frac{1-\delta^n}{1-\delta}.$$

Given that $\lambda > 0$ and $0 < \delta < 1$, we have that $\lambda_n > 0$ and $\lambda_{n+1} > \lambda_n$ for every n in S. This implies that the network search intensity is an increasing function of the number of immigrants. Consequently, every addition to the network reduces the expected time between job observations.[9]

It is worth pointing out that while there is no interaction or information transmission between the immigrants in the network, their combined search efforts reduce the expected observation interarrival times. We call this effect increase of the job search efficiency.

2.1.2 Reduction of vacancies

Thus far we have that as the network grows, the expected frequency of job observation increases. The crucial fact here is that under the assumptions of fixed labor demand and no exit from employment, the expansion also reduces the number of vancancies. This implies that the network should make -on average- more observations to find an available job.

Suppose that the size of the network is n, for some n in S. Then conditional on making an observation, the probability of selecting a vacancy is $p_n = \frac{N-n}{N-1}$. Note that as long as the network size remains the same, every job observation is a Bernoulli trial with success probability p_n. Therefore, we can think of the number of observations before selecting a vacant job as a geometric variable of parameter p_n, with support on $\{1, 2, \ldots\}$. It follows that the average number of observations in which a network of size n finds a vacancy is $1/p_n$. One sees immediately that an increase of the network size enlarges the average number of observations needed to find a vacant job.

[9] If we were to assume the number of immigrants in the network is a continuous variable, we would have $\frac{\partial^2 \lambda_n}{\partial n^2} = \frac{\lambda \delta^n log^2(\delta)}{\delta - 1} < 0$. This implies that the marginal increase in efficiency is smaller as the network expands. Evidently, this is due to the decreasing nature of the immigrant search intensities.

It is worth pointing out that once an available job is found and in consequence, an immigrant is added to the network, both the network search efficiency and the probability of success adjust to the new network size. Hence, the principal significance of the reduction of vacancies is that it induces a *thinning*[10] of the process of observations. In other words, the expansion of the network induces a higher intensity search process in which successful observations occur with a lower probability. This amounts to saying that the search process of the network is a Poisson process whose rates vary according to the counting process of immigrants.[11]

2.1.3 Interaction between search efficiency and decreasing vacancies

In exploring the events that lead to network expansion, we noted that the immigrant interarrival times are affected by the number of job observations in which the network selects a vacancy and the frequency of these observations. What is still lacking is an explicit description of the distribution of the interarrival times.

We begin with the observation that the interarrival times of immigrants are given by the time in which the network observes an available job. For the general case, set $X(t) = n$, for some n in S. Let T_n denote the time that elapses before one of the n immigrants in the network makes the first observation of a vacant job and invites a potential migrant to join the network. Another way of stating this is to say: T_n is the time spent by $\{X(t)\}_{t\geq 0}$ in state n before jumping to state $n+1$. We refer to T_n as the holding time in state n, or the interarrival time of the $(n+1)$-th immigrant.

Proposition 1. *The holding time T_n has an exponential distribution of parameter $q_n = \lambda_n p_n$, where $\lambda_n = \lambda \frac{1-\delta^n}{1-\delta}$ is the network search intensity, and $p_n = \frac{N-n}{N-1}$ is the probability of observing a vacancy when the network size is n.*[12]

[10] See Grimmett and Stirzaker, 2001

[11] See Scott and Smyth (2003) for the definition of Markov Modulated Poisson Process.

[12] For the proof we refer the reader Appendix A.

Therefore, conditional on $X(0) = 1$, the holding times $T_1, T_2, \ldots, T_N$, are independent[13] exponential random variables of respective parameters $q_1, q_2, \ldots, q_N$. This actually proves that the counting process of immigrants is a birth process of rates $(q_n)_{n\in S}$.[14]

It is worth pointing out that since the probability of observing a vacancy when $X(t) = N$ is zero, we conclude from Proposition 1 that $q_N = 0$, hence that $T_N \sim \exp(0)$, and finally that $\mathbb{E}(T_N) = \infty$. This means that the network size spends an infinite time in state N. Thus, N is an absorbing state of the counting process of immigrants. This is consistent with the fact that once there are N immigrants in the destination, no job may be offered to the $(N+1)$th immigrant, and so the network size remains unchanged.

For a deeper discussion of the behavior of the birth rates (or network search intensities), we exemplify the effect of a change in the network search effiency -through the parameter of search intensity reduction δ- on the overall intensities $(q_n : 1 \leq n \leq N)$.

Example 1. *We set a maximum network size of $N = 10$. Since λ is a scale parameter, there is no loss of generality in assuming $\lambda = 1$. For this exercise we consider four cases of intensity reduction, $\delta \in \{0.01, 0.50, 0.75, 0.99\}$. The first case accounts for a low network search efficiency, since in the limiting case $\delta = 0$ only the pioneer immigrant searches for jobs.[15] The last case portrays a high efficiency of search, given that in the limit $\delta = 1$, all the network members search at the same full search intensity λ. The second and third cases represent intermediate search efficiencies.*

Figure 1 shows the network search intensities that correspond to each value of δ and network size n in $\{1, \ldots, 10\}$. In the plot, low values of δ and therefore low network intensities have light colored graphs, and the intensity of color increases as δ grows. The rates are joined by lines to exhibit their trend as the network expands. It can be seen that when $\delta = 0.01$, the

[13]Independence follows from the Markov property of the immigrant counting process: $\mathbb{P}\left(X(T_n) = n+1 \middle| X(T_{n-1}) = n, \ldots, X(0) = 1\right) = \mathbb{P}\left(X(T_n) = n+1 \middle| X(T_{n-1}) = n\right)$.

[14]See Norris (1998).

[15]We adhere to the convention that $0^0 = 1$, so when $\delta = 0$ the search intensities are $\lambda_1 = \lambda$ for the first immigrand, and $\lambda_n = 0$ for every $1 < n \leq N$.

network search intensities appear to be linear. This is because as δ approaches zero we have

$$\lim_{\delta \to 0} q_n = \frac{\lambda}{N-1}(N-n) \quad \textit{for every } n \in S. \tag{1}$$

Clearly, when the network search efficiency is low, the prevailing effect in the birth rates is that of the increased occupation of jobs. This explains the decreasing behavior of the overall search intensities as the network expands. Since the interarrival times of immigrants have exponential distribution of parameters $(q_n)_{n \in S}$ and consequently, expected interarrival times $\mathbb{E}(T_n) = \frac{1}{q_n}$, we see that every addition to the network increases the average times in which subsequent immigrants arrive.

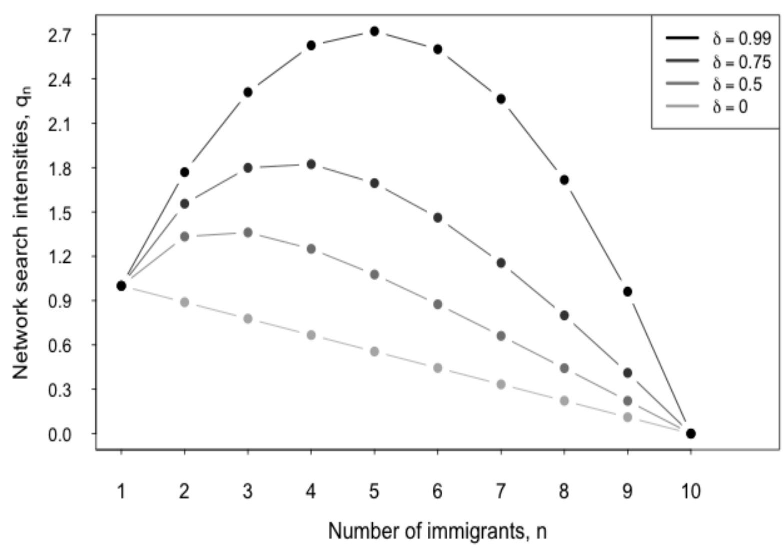

Figure 1: Network search rates by search efficiency and network size

On the other hand, as δ approaches 1, the birth intensities take the form

$$\lim_{\delta \to 1} q_n = \frac{\lambda}{N-1} n(N-n). \tag{2}$$

In this case, the rates are symmetric around the network size $\frac{N}{2}$ *if* N *is even, or* $\frac{N+1}{2}$ *if* N *is odd. This means that the average time in which* n *immigrants make a successful job observation is equal to that of* $N-n$ *immigrants.*[16] *The reason for this is that when* n *immigrants look for jobs, the expected times between observations are longer than when there are* $N-n$ *members, but the probability of success is higher, and this counteracts the small search efficiency effect. On the other hand, when* $N-n$ *immigrants search, the overall search efficiency is higher than when there are* n *immigrants, but the probability of selecting one of the* n *vacancies is small. From this we deduce that the net effect is the same for both network sizes.*

We now consider the behavior of the birth rates for any other δ *in* $(0,1)$*. The important point to note here is that sequential increases in the network search efficiency take the birth rates function from a decreasing 'line' to a discrete parabola that opens down, as seen in Figure 1. Our proof starts with the observation that for any* δ *in* $(0,1)$*, we have* $q_1(\delta)=1$ *and* $q_N(\delta)=0$*, so the birth rates functions have the same start and end points. To study the rest of the cases, take any* δ_0 *and* δ_1 *in* $(0,1)$ *such that* $\delta_0<\delta_1$*, and note that for every state* $n\in S$ *we have* $\delta_0^n<\delta_1^n$*. Given that* $\lambda>0$*, it follows easily that* $\lambda_n(\delta_0)<\lambda_n(\delta_1)$*. As* p_n *is not affected by* δ*, this clearly forces*

$$q_n(\delta_0) < q_n(\delta_1) \quad \textit{for every } n \textit{ in } S. \tag{3}$$

It follows that the network seach intensities are increasing in δ*, and thus in the network search efficiency. This can be confirmed graphically in Figure 1, where we see that* $q_n(0.5) < q_n(0.75)$ *for every* $1<n<10$*.*

We emphasize that the birth rates associated to δ *in* $\{0.5, 0.75, 0.99\}$ *have an initial tendency to increase, reach an stationary point and decrease from that point on. This suggests that the network search efficiency overpowers the reduction of job opportunities at first, but once the network reaches a critical size, the dominance of effects acts in reversed order.*

[16] Without loss of generality we are assuming N is even.

As the birth rates functions are increasing in the search intensity, we have that the critical network sizes go from one (the size where the line given in Equation (1) *reaches its local maximum), to* $\frac{N}{2}$ *(where Equation* (2) *has its critical point). The interest of the stationary point is that it allows one to determine the dominating effect in the network search rates, we will see however, that this critical size may not correspond to the economic saturation point of the network.*

Summarizing, we have established the evolution of a labor migration network, from the arrival of its first member until the maximum size is reached. We worked under the assumption that every potential immigrant that is invited to join the network does it. We know this is not necessarily true, however the assumption allowed us to derive the state transitions of $\{X(t)\}_{t\geq 0}$ and its corresponding birth rates $(q_n)_{n\in S}$. In the following section, we show how to derive the expected returns to network-driven migration from the rates we obtained.

3 Expected returns

Recall that the allocation of a potential migrant is the result of her optimal choice when comparing the expected values of staying in the home country, migrating to a location with a network and starting a network elsewhere. This comparison will allow us to determine whether joining a network of size $n \in S$ is the optimal strategy of a worker. Needles to say, this will let us dispense with the assumption of potential immigrants joining the network unconditionally. This basic idea will also allow us to make an economic formulation of the network saturation point.

In this section we define the expected returns of the options available to potential migrants: stay in the origin, migrate to a new location in the receiving country and look for a job in a destination through a network of size $n \in S$. We follow Todaro (1969) in assuming that: the planning horizon for potential migrants is identical and infinite, there is an initial fixed cost of migration and relocation for the first immigrant c_a, and the probability of employment in the home country is one.

In addition, we make a few assumptions trying to keep the model as simple as possible. The sending country has a relative low wage w_s when compared to the host wage $w > w_s$. The underlying assumption is that the labor supply shifts due to migration are small enough that their impact on the source and destination wages is inconsequential.[17] Clearly, this assumption allows us to do only a partial-equilibrium analysis. We also assume that the real instantaneous rates of interest at origin r_s and destination r are constant and exogenous. In addition, we consider that there are costs involved with the use of networks to search for jobs.

Let V_s denote the payoff of staying in the origin country, and h be the length of a period. Since workers are employed with probability one in the origin, it is easily seen that

$$V_s = \frac{1}{1 + r_s h}[w_s h + V_s].$$

This implies:

$$V_s = \frac{w_s}{r_s}. \tag{4}$$

We define V_a to be the expected return to start a network in a host destination. Since the host wage is the same in all locations and the cost of starting a network does not depend on the migration distance, we have that the payoff V_a is the same in all the destinations. Given that the pioneer immigrant secures one of the N jobs in the host market with probability one, and pays the fixed cost c_a it follows easily that

$$V_a = \int_0^\infty w e^{-rt} dt - c_a.$$

[17]Since we are interested in modeling the effect of the reduction of vacancies -rather than the effect of the wage distribution- on the migrating decision, the choice of constant wages seems to be the best adapted to our theory. However, as Borjas (2014) points out, a general equilibrium framework requires to endogenize the parameters of the income distributions and allow them to depend on the size of the migration flows.

This yields

$$V_a = \frac{w}{r} - c_a. \tag{5}$$

Condition to migrate

Let us orient the decision to migrate by the requirement that the expected return differential between starting a network and staying in the origin be non-negative. Then, by Equations (4) and (5) international migration occurs if

$$c_a \leq \frac{w}{r} - \frac{w_s}{r_s}. \tag{6}$$

This constraint is unambiguous as long as $\frac{w_s}{r_s} < \frac{w}{r}$. Geometrically speaking, (6) ensures that the pioneer's expected return V_a is above (or the same as) V_s.

Expected return to use a network

It is our interest to see if the use of networks to search for jobs increases the expected gains of aspiring immigrants. The basic idea is that workers pay the network's search cost (c_n for a network of size n in S) and get the present value of income from search. Two results may arise: either the network finds a vacancy for the worker -in which case she takes the job and joins the network-, or no available job is observed -in which case the worker stays in the source country and decides whether or not to engage in search again-. In terms of the counting process of immigrants $\{X(t)\}_{t\geq 0}$, the search results in the addition of one or zero members to the network.[18] In what follows, we determine the contribution of each of these events to the expected return of network-driven migration.

For the general case, set $X(t) = n$. By the *infinitesimal definition* of $\{X(t)\}_{t\geq 0}$,[19] the probability of transitioning from a network size n to a size $m \in S$ in a short interval of time h is given by:

[18] We will prove that the likelihood of having two or more additions in a small time frame is negligible.

[19] See Norris, 1998.

$$p_{nm}(h) = \begin{cases} q_n h + o(h) & \text{if } m = n+1, \\ 1 - q_n h + o(h) & \text{if } m = n, \end{cases} \tag{7}$$

where the term $o(h)$ is a function of h such that $\lim_{h \to 0} o(h) = 0$.

We prove the first equation to shed light on the infinitesimal definition and its relation with the birth rates $(q_n)_{n \in S}$. We are interested in computing the probability that at time $t + h$ the network size is $n + 1$ given that $X(t) = n$. Let us first observe that the distribution of the network size $X(t+h)$ conditionally on $X(t)$ does not depend on t, so

$$\begin{aligned} \mathbb{P}(X(t+h) \geq n+1 \mid X(t) = n) &= \mathbb{P}(X(h) \geq n+1 \mid X(0) = n) \\ &= \mathbb{P}(T_n \leq h \mid X(0) = n). \end{aligned}$$

As the holding time T_n has an exponential distribution of parameter q_n, we have $\mathbb{P}(T_n \leq h \mid X(0) = n) = 1 - e^{-q_n h} = q_n h + o(h)$, so

$$\mathbb{P}(X(t+h) \geq n+1 \mid X(t) = n) = q_n h + o(h). \tag{8}$$

One may prove by a similar argument that $\mathbb{P}(X(t+h) \geq n+2 \mid X(t) = n) = o(h)$, which implies that the probability of adding two or more members in a small time frame is negligible. It suffices to use this observation together with Equation (8) to get the desired result.

Equation (7) may be summarized by saying that the network members find a vacancy with probability $q_n h + o(h)$. In this event, the aspiring immigrant gets the job and earns the host wage w forever. On the other hand, it may happen that the network members make no job observations in the time interval $(t, t+h)$, or that one of them observes a filled job. These events occur with total probability $1 - q_n h + o(h)$. In this case, the cardinality of the network remains the same and consequently, so does the expected return to use the network. It follows that the expected return to search through a network of size n, V_n, is given by

$$V_n = \frac{-c_n h}{1 + rh} + \frac{q_n h}{1 + rh} \int_h^{\infty} w e^{-rx} dx + \frac{1 - q_n h}{1 + rh} V_n + o(h), \tag{9}$$

where c_n is the network search cost function. The classical literature on network driven migration suggests that the expansion of networks lowers the costs of migration (Massey, 1990; Massey *et al.*, 1994.), so the cost function is so chosen that $c_n \leq c_{n-1}$ for all $n > 1$ in S.

It is worth stressing an important difference between Equation (9) and standard job search Bellman equations. It is derived from the fact that the host wage distribution $F(x)$ is deterministic,[20] so the search that the network carries is not aimed at finding an 'acceptable' job, but simply at finding *a* job. This is the reason why the payoff of a job offer is not a random variable of the form $\int_0^\infty \max\left\{V_n, \frac{x}{r}\right\} dF(x)$, but simply the infinite accumulation of the host wage w.

Rearranging terms in (9) and taking limit as $h \to 0$ yields:

$$V_n = \frac{-c_n}{r+q_n} + \frac{q_n}{r+q_n}\left(\frac{w}{r}\right). \tag{10}$$

It is clear that the expected value of using a network is decreasing in the search costs and the interest rate, and increasing in the host wage. However, the relation of V_n with respect to the search intensities is not evident given the non-monotonic behavior of the birth rates function q_n.

Condition to use networks

We assume that a potential immigrant searches through a network of size n if the expected return to do so is greater than the expected return to start a network elsewhere. We emphasize that there is no guarantee that for any non-decreasing cost function c_n, the expected returns of network-driven migration are higher than the expected return to start a network. In the extreme case, c_n may be such that $V_n < V_a$ for every n in S. Geometrically speaking, this would imply that the network value function is always below the line V_a. Needless to

[20] By assumption, the wage distribution function $F(x)$ takes the value 0 if $x < w$, and 1 if $x \geq w$.

say, this means that there are no incentives to use the network. Clearly, we must impose an additional condition on c_n to ensure that networks are in fact used.

Under the assumption stated above, a necessary and sufficient condition for network use is that $V_a < V_1$.[21] Thus, the requirement on c_1 is that:

$$c_a r - c_1 > w - \lambda c_a. \tag{11}$$

This is nothing but the statement that networks of size one are used when the (per period) opportunity cost of being a pioneer $c_a r - c_1$ is greater than the migration payoff w plus the expected gain of starting a network, which in this case is the probability that the network finds a job λ, times the expected increase in value associated with starting a network $-c_a$.

Constraints (6) and (11) combined give $c_1 < \lambda(\frac{w}{r} - \frac{w_s}{r_s}) - w_s \frac{r}{r_s}$.

4 Economic saturation of networks

Different authors have proposed definitions of network saturation, either in terms of the migration costs (Massey *et al.*, 1990), the earnings of immigrants in a destination relative to earnings in the country (Heer, 2002), or the number of jobs for immigrants in a given destination (Light, 2006). These measures place saturation in a classic supply-demand framework in which the expansion of networks increases the immigrant labor supply and decreases the wages/jobs of subsequent immigrants. What is still lacking is an analytic model of this interaction and an explicit description of the network saturation point. The aim of this paper is to provide an economic criterion of network saturation that accounts for the increase in the job search efficiency and the reduction of vacancies that the expansion of networks brings about.

Definition 1. *We define the network saturation point to be the smallest network size for which the value of using the network is smaller than the value of starting a new one. In*

[21] Note that if this does not hold, immigrants will create networks but not use them.

other words, the saturation point gives the maximum network size for which network-driven migration is optimal. In our notation, this critical network size n^* *is given by*

$$n^* = inf\{n \in S : V_n \leq V_a\}. \tag{12}$$

Although it would be desirable to obtain an explicit formula of the saturation point, we have not been able to do this because the network search intensities $(q_n)_{n\in S}$ are not analytically tractable. Nonetheless, we find that the saturation point can be uniquely determined once the parameters of the model are set. We give a few examples below.

Example 2. *Consider the previous example in which* $N = 10$, $\lambda = 1$ *and* δ *in* $\{0.01, 0.5, 0.75, 0.99\}$. *For the instantaneous origin and destination wages we use Albrecht et al. (2009) Latin American average informal and formal-sector wages. The crucial assumption is that immigrants in the origin get the informal-sector wage* $w_s = \$0.20$*, and a wage in the destination country that is equivalent to the formal-sector wage* $w = \$0.40$.[22] *For simplicity we assume that* $r = r_s$ *and set* $r = 0.015$. *We set* $c_a = \$10$ *and* $c_n = \$5$ *for every* n *in* $\{1, \ldots, 10\}$. *It is evident that the chosen costs satisfy conditions* (6) *and* (11).

Figure 2 pictures the expected network values generated with these parameters. We adhere to the convention that the depth of color reflects the network's intensity of search, so the darkest graph is associated to $\delta = 0.99$ *and the lightest to* $\delta = 0.01$. *The expected value of staying in the origin is given by the dotted line (*$V_s = \$13$*), while the straight line gives the expected return to start a network (*$V_a = \$17$*). We see that the expected returns to network-driven migration are increasing in the network's search intensity.*[23] *From this we deduce that higher network search intensities give greater expected returns and in consequence, bigger saturation sizes. It is worth stressing that aspiring immigrants take the network size and*

[22]The average formal-sector wage used by Albrecht (2009) is 0.363.

[23]We prove this result in general. We conclude from (3) that $0 < \delta_0 < \delta_1 < 1$ implies $q_n(\delta_0) \leq q_n(\delta_1)$ for every $1 \leq n < N - 1$, hence that $\frac{q_n(\delta_0)}{r+q_n(\delta_0)} \leq \frac{q_n(\delta_1)}{r+q_n(\delta_1)}$, and finally that $V_n(\delta_0) \leq V_n(\delta_1)$. The strict equality holds for $n = 1$.

search intensity as given and decide whether using the network maximizes their discounted value of income stream.

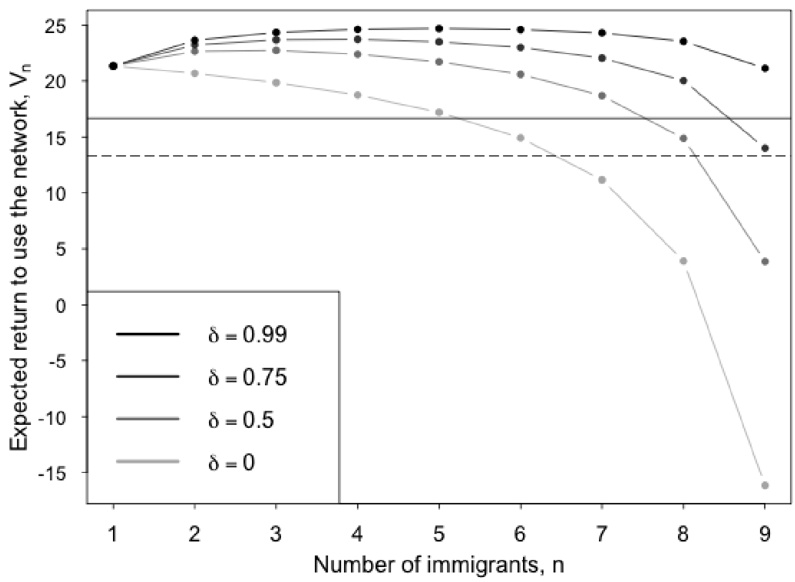

Figure 2: Expected return of network-driven migration.

Note that if the cost of starting a network were to decrease, ceteris paribus, there would be an upward shift of the pioneer's expected return line. Consequently, the critical size of the network would be reduced. The obvious intuitive meaning is that a decrease in c_a *makes the option of starting a network more appealing to potential immigrants.*

Figure 2 points out another interesting aspect of the expected returns to using a network: their behavior is similar to that of the network intensities. We see that in the early stages of network evolution, the expected returns associated with δ *in* $\{0.50, 0.75, 0.99\}$ *have a tendency to increase, reach a maximum and decrease. We infer that this is the result of network search efficiency overpowering resource reduction first, and then overturning.*[24] *As it is expected,*

[24]A similar result is found when the search costs account for the search efficiency, so $c_n = c\lambda_n$ for every

when $\delta = 0.01$, the average returns to network migration decrease with every immigrant joining the network. This follows from the network's lack of search efficiency and the assumption of constant search costs.

Table 1 presents the expected returns to network-driven migration by network size and value of δ. The saturation values that correspond to each case, $V_{n^*}(\delta)$, are marked with an asterisk. One sees that the saturation point is not necessarily equal to the maximum size of the network $N = 10$. In our example, the saturation sizes for δ in $\{0.01, 0.5, 0.75, 0.99\}$ are 6, 8, 9 and 10 respectively. This result shows that the creation of networks in new locations may occur before the targeted destination runs out of vacancies. Incidentally, we have shown that it is not necessary to account for the reduction of wages to observe an early deflection of migration. It suffices to consider the diminishment of jobs for immigrants and its negative effect on the network search intensities. We emphasize that this outcome does not neglect the increasing (geometric) efficiency of search accomplished by the network members. This is our main result.

We claim that the immigration of a single worker to the host country is a sufficient condition for the proliferation of networks. This is because the arrival of an immigrant implies that starting a network is revealed preferred to staying in the source. If condition (11) holds, the pioneer's network will be used by aspiring immigrants until it becomes saturated. In this point, the remaining allocation strategies are staying in the source and migrating to a new location. It is optimal for workers to start a network since $V_s \leq V_a$. It remains to prove that if (11) does not hold, networks will be created constantly. We conclude that $V_1 > V_a$, and given that $V_a > V_s$, networks will always be created but not used. This completes the proof.

---

n in S, and c a positive constant. This case contradicts the network theory assumption of falling costs and this is our reason for excluding it from the main analysis.

Table 1: Expected returns to network-driven migration by network size and search intensity

	$\boldsymbol{\delta}$			
n	0.01	0.50	0.75	0.99
1	21.35	21.35	21.35	21.35
2	20.75	22.66	23.23	23.64
3	19.92	22.74	23.69	24.34
4	18.82	22.40	23.73	24.62
5	17.29	21.72	23.51	24.69
6	15.03*	20.60	23.01	24.60
7	11.31	18.68	22.05	24.30
8	4.12	14.87*	20.04	23.55
9	-15.78	3.86	13.99*	21.13
10	-333.33	-333.33	-333.33	-333.33*

Note: The saturation values $V_n^* = \inf\{V_n : V_n \leq V_a = 17\}$ are marked with an asterisk.

4.0.1 Network saturation time

Although literature on network migration provides a general description of the conditions for network saturation, rather less attention has been paid to the time in which networks become saturated. This is to be expected, since the theory of network migration only assumes that networks find jobs for potential members, yet it does not explicitly describe the times in which such jobs are found. The advantage of our model lies in the fact that it provides a natural characterization of the immigrant interarrival times. This feature allows us to shed new light on the economic saturation time of a network.

We begin by noting that the time in which a network reaches its saturation point is given by the accumulation of the interarrival times of the n^*-1 immigrants arrived after the pioneer. Since these interarrival times are exponential variables of parameters $q_1, \ldots, q_{n^*-1}$, it follows that the expected saturation time is $\sum_{n=1}^{n^*-1} \mathbb{E}(T_n) = \sum_{n=1}^{n^*-1} \frac{1}{q_n}$. Although this average time is descriptive, it is of interest to know the actual distribution of the economic saturation time of the network. The main difficulty in this task is that the interarrival times of immigrants are not identically distributed.

Instead of computing the characteristic or moment generating function of the saturation time we use some results of phase-type distributions, a class of distributions of the time in which some Markov processes (like our finite-state birth process) reach their absorbing state. We will touch only a few aspects of the theory trying to keep our exposition self-contained.[25]

Let τ be the time it takes the network reach its absorbing state,

$$\tau := \inf\{t : X(t) = N\}.^{26} \tag{13}$$

By definition, τ denotes the time until complete network saturation, so we also have

$$\tau = W_N = \sum_{i=1}^{N-1} T_i,$$

[25] See Mogens (2005) for a general reference of phase-type distributions.

[26] τ is also known as the time until absorption.

where the holding times $T_1, \ldots, T_N$ are exponential random variables with respective parameters $q_1, \ldots, q_N$.

Proposition 2. *The density function of τ is given by*

$$f_\tau(x) = \sum_{i=1}^{N-1} q_i e^{-q_i x} \left(\prod_{j=1, j\neq i}^{N-1} \frac{q_j}{q_j - q_i} \right).^{27} \tag{14}$$

This proposition states that the complete saturation time has a hypoexponential distribution of parameters $\{q_n\}_{n=\{1,\ldots,N-1\}}$.

Proposition 2 gains in interest if we realize it also allows us to compute the economic saturation time of networks. It suffices to note that saturated networks are not used by potential immigrants. Explicitly, this means that once a network reaches the critical size n^*, its cardinality will not change. This allows us to think of n^* as an *economic absorbing state* of $\{X(t)\}_{t\geq 0}$. By the above, we can think of $N = n^*$ to be the 'new' maximum size. It is immediate that the network's time until economic absorption is given by

$$\tau^* = \inf\{t : X(t) = n^*\}.$$

From Proposition 2 we conclude that τ^* is a hypoexponential random variable with parameters $\{q_n\}_{n=\{1,\ldots,n^*-1\}}$.[28]

This section was intended as an attempt to motivate the study of network saturation in the presence of frictions. Central to our viewpoint is the notion that the network must engage in search to find vacancies, and that the search itself takes time. There has since been little systematic work on migration network models with frictions, yet the time approach is not much different from (unemployment) duration analysis with time-varying hazard rates.

[27] For the proof we refer the reader to in Appendix C.

[28] The distribution of τ^* is given by $F_{\tau^*}(x) = p_{1,n^*}(x)$. See Appendix C for the proof.

5 Efficiency

One may ask whether it is efficient for the immigrant population to reach the saturation point of networks. To answer this we next introduce the notion of expected gains of network migration. Let $V_n - V_a$ denote the expected gain that immigrant $n+1$ gets from using a network of size n rather than starting one on her own, for $1 \leq n < N$. Clearly, the expected gains are zero or negative if i is in $\{n^*, \ldots, N-1\}$. We define the expected gain of a network to be the sum of the expected gains of all the network members except the pioneer,[29] and the expected gains of network migration to be the sum of the expected gains of all the networks in the destination country. The task is now to determine whether reaching the saturation point of networks n^* maximizes the expected gains of network migration.

For simplicity we assume that there is only one network in the host contry and that it has size n^*. The expected gain of the network is then

$$\sum_{n=1}^{n^*-1} (V_n - V_a). \tag{15}$$

To prove that reaching the saturation point is not efficient amounts to finding a different arrangement of n^* immigrants in more networks that yields a higher expected gain than this. For instance, there could be an arrangement such that

$$\sum_{n=1}^{n^*-1} (V_n - V_a) < \sum_{n=1}^{k} (V_n - V_a) + \sum_{n=1}^{n^*-k-2} (V_n - V_a), \tag{16}$$

where the right side of the inequality gives the expected gain when there are two networks: one of size $k+1$ and one of size $n^* - k - 1$. If no such arrangement exists, we say it is efficient to reach the saturation point.

Note that the number of networks in the new arrangement is constrained by the saturation point n^* and the fact that pioneer immigrants have no gains of network migration.

[29] It is obvious that the pioneer does not rely on networks to migrate, therefore she has no expected gains of network migration.

Example 3. *Consider the case in Table 1 where* $\delta = 0.5$*. The saturation point associated to this parameter is* $n^* = 8$ *and its corresponding network's expected gain is* $\sum_{n=1}^{7}(V_n - 17) =$ \$33.48*. Note that if instead of having a network of size 8, workers form two networks of size 4, the expected gain of network migration becomes* $2\sum_{n=1}^{3}(V_n - 17) =$ \$33.50*. This means that the value* $k = 3$ *satisfies* (16)*. We conclude that in this case, it is more efficient for immigrants to create a new network after reaching the highest network expected gain* $V_3 - V_a = 6.08$*, than to wait until the saturation point* $n^* = 8$ *is reached.*

This example demonstrates rather strikingly that it may be more efficient for immigrants as a whole to deflect migration before the model predicts it should.

6 Conclusions

The aim of this paper is to propose a tractable definition of the saturation point of labor migration networks. For this purpose we constructed a stochastic process that models the evolution of migrant networks. The model provides an intrinsic characterization of the expansion of networks when its members have incomplete information about the location of vacant jobs. This construction relaxes the standard assumption on social network theory that networks instantaneously find jobs for aspiring immigrants.

The main idea is that workers in the origin may use networks to search for jobs in host locations. Our basic assumption is that aspiring immigrants pay the search costs incurred by the network and get the present value of income from search. The behavior of the network search intensities reflects the fact that as networks grow, the job search efficiency increases and vacancies in the targeted destination decrease. Accordingly, the network samples jobs faster (on average), but takes longer to select a vacant job (also on average). The expected return to search through a network is computed by means of a Bellman equation.

We are interested in finding the size of networks for which the expected return to network migration is lower than the average return of migrating by oneself. This criterion allows to express the saturation point as the maximum size for which the use of networks is an optimal strategy. We give a few examples that demonstrate that the saturation point does not necessarily correspond to the maximum size of the network. This important result suggests that the use of networks is not always preferred by migrants. We also show that in some cases, the average gains of network-driven migration may be increased by creating networks of smaller size than the one predicted by the model. This implies that an early deflection of migration to new destinations could be more efficient for the immigrant population as a whole.

Lastly, we discuss the network saturation time. To our knowledge of the literature, no attempt has been made to develop a theory of the time in which networks become saturated. We find that under the above assumptions, the saturation time of networks is a hypoexponential random variable with parameters given by the network search intensities.

References

[1] Albrecht, James, Lucas Navarro, and Susan Vroman. 2009. The effects of labour market policies in an economy with an informal sector. *The Economic Journal* 119 (July): 1105-1129.

[2] Borjas, George J. 2014. *Immigration economics.* Cambridge: Harvard University Press.

[3] Caliendo, Marco, Ricarda Schmidl, and Arne Uhlendorff. 2010. Social Networks, Job Search Methods and Reservation Wages: Evidence from Germany. IZA Discussion Paper no. 5165, Institute for the Study of Labor.

[4] Calvó-Armengol, Antoni, and Matthew O. Jackson. 2007. Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory* 132: 27-46.

[5] Corless, Robert M., Gaston H. Gonnet, David EG Hare, David J. Jeffrey, and Donald E. Knuth. 1996. On the LambertW function. *Advances in Computational mathematics* 5, no. 1: 329-359.

[6] Durand, Jorge, and Douglas S. Massey. 2003. *Clandestinos: Migración México-Estados Unidos en los albores del siglo XXI*. México: Miguel Ángel Porrúa and Universidad Autónoma de Zacatecas.

[7] Fussell, Elizabeth, and Douglas S. Massey. 2004. The Limits to Cumulative Causation: International Migration from Mexican Urban Areas. *Demography* 41, no. 1: 151-171.

[8] Grimmett, Geoffrey R., and David R. Stirzaker. 2001. *Probability and Random Processes*. New York: Oxford University Press.

[9] Heer, David M. 2002. When cumulative causation conflicts with relative economic opportunity: Recent change in the Hispanic population of the United States. *Migraciones Internacionales* 1, no. 3: 32-53.

[10] Kingman, J.F.C. 1993. *Poisson Processes*. New York: Oxford University Press.

[11] Light, Ivan. 2006. *Deflecting Immigration: Networks, Markets, and Regulation in Los Angeles*. New York: Russell Sage Foundation.

[12] Massey, Douglas S. 1990. Social structure, household strategies, and the cumulative causation of migration. *Population index*: 3-26.

[13] Massey, Douglas S., Rafael Alarcón, Jorge Durand, and Humberto Gonzalez. 1990. *Return to Aztlan: The social process of international migration from Western Mexico*. Berkeley: University of California Press.

[14] Massey, Douglas S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J. Edward Taylor. 1994. An evaluation of international migration theory: The North American case. *Population and development Review*: 699-751.

[15] Massey, Douglas S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J. Edward Taylor. 1998. *Worlds in motion: Understanding International Migration at the End of the Millennium.* Oxford: Clarendon Press.

[16] McKenzie, David, and Hillel Rapoport. 2010. Self-selection patterns in Mexico-US migration: the role of migration networks. *The Review of Economics and Statistics* 92, no. 4: 811-821.

[17] Mogens, Bladt. 2005. *A review on phase-type distributions and their use in risk theory.* Astin Bulletin 35, no. 1: 145-161.

[18] Monras, Joan. 2015. Immigration and wage dynamics: Evidence from the mexican peso crisis. IZA Discussion Paper no. 8924, Institute for the Study of Labor.

[19] Myrdal, Gunnar. 1957. *Rich lands and poor: the road to world prosperity.* New York: Harper & Row.

[20] Norris, James R. 1998. *Markov chains.* Cambridge University Press.

[21] Scott, Steven L., and Padhraic Smyth. 2003. The Markov Modulated Poisson Process and Markov Poisson Cascade with Applications to Web Traffic Modeling. *Bayesian Statistics* 7: 671-680.

[22] Taylor, J. Edward. 1986. Differential migration, networks, information and risk. In *Research in Human Capital and Development, vol. 4, Migration, Human Capital, and Development*, ed. Oded Stark.

[23] Todaro, Michael P. 1969. A model of labor migration and urban unemployment in less developed countries. *The American economic review* 59, no. 1: 138-148.

[24] Todaro, Michael P. 1980. Internal migration in developing countries: a survey. In *Population and economic change in developing countries*, ed. Richard A. Easterlin.

Appendix A Holding times

Proposition 3. *Suppose that $X(t) = n$ for some n in S and $t \geq 0$. Then, the holding time T_n has exponential distribution with parameter $\lambda_n p_n$, where $\lambda_n = \lambda \frac{1-\delta^n}{1-\delta}$ and $p_n = \frac{N-n}{N-1}$.*

Our proof starts with the observation that when the cardinality of the network is n, the rate of the network search process is $\lambda_n = \frac{\lambda(1-\delta^n)}{1-\delta}$. This means that as long as the number of immigrants in the network remains the same, the interarrival times of observations denoted by $S_1, S_2, \ldots$ are independent exponential variables of parameter λ_n.

Let K stand for the number of observations made by the network members until one of them observes an available job. Here, it is natural to assume K is independent of the sequence $(S_i)_{i=1}$. Under the assumption that there are n immigrants employed at destination, the probability of making a successful observation is given by $p_n = \frac{N-n}{N-1}$. We thus claim that K is a geometric random variable of parameter p_n.

The holding time in state n, is given by

$$T_n = \sum_{i=1}^{K} S_i. \tag{17}$$

To deduce the behavior of the distribution of T_n, we proceed to compute its characteristic function.

Let us first condition on K to obtain

$$\begin{aligned}
\Phi_{T_n}(t) &= \mathbb{E}\left[e^{itT_n}\right] \\
&= \mathbb{E}\left[\exp\left\{it\sum_{i=1}^{K} S_i\right\}\right] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left(\exp\left\{it\sum_{i=1}^{k} S_i \middle| K=k\right\}\right)\mathbb{P}(K=k) \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left(\exp\left\{it\sum_{i=1}^{k} S_i\right\}\right)\mathbb{P}(K=k)
\end{aligned}$$

The last equality holds by independence of K and the sequence of interarrival times $(S_i)_{i=1}$.

Given that $S_1,\dots,S_k$ are independent exponential variables of parameter λ_n, it follows easily that $\mathbb{E}\left(\exp\left\{it\sum_{i=1}^{k} S_i\right\}\right)$ is the characteristic function of a Gamma variable of parameters k and λ_n valued at t. This gives

$$\Phi_{T_n}(t) = \sum_{k=0}^{\infty}\left(\frac{\lambda_n}{\lambda_n - it}\right)^k \mathbb{P}(K=k) = \mathbb{E}\left[\left(\frac{\lambda_n}{\lambda_n - it}\right)^K\right]. \tag{18}$$

We can rewrite $\left(\frac{\lambda_n}{\lambda_n-it}\right)^K$ as $e^{iK\frac{1}{i}\log\left(\frac{\lambda_n}{\lambda_n-it}\right)}$, and make the variable change $u=\frac{1}{i}\log\left(\frac{\lambda_n}{\lambda_n-it}\right)$, so Equation (18) can be expresed as

$$\Phi_{T_n}(t) = \mathbb{E}\left[e^{iuK}\right]. \tag{19}$$

Note that by definition of K, $\mathbb{E}\left[e^{iuK}\right]$ is the characteristic function of a geometric random variable of parameter p_n valued at u. Now Equation (19) becomes:

$$\Phi_{T_n}(t) = \frac{p_n e^{iu}}{1-(1-p_n)e^{iu}}. \tag{20}$$

Substituting the value of u on equation (20) yields

$$\Phi_{T_n}(t) = \frac{\lambda_n p_n}{\lambda_n p_n - it}. \tag{21}$$

It is clear that this is the characteristic function of an exponential random variable of parameter $q_n = \lambda_n p_n$ valued at t. This is the desired conclusion.

Appendix B Markov semigroup

The forward and backward systems of equations of a birth process with finite state space $S = \{1, \ldots, N\}$ and intensities $\{q_1, \ldots, q_N\}$, are respectively:

$$p'_{nm}(t) = p_{n,m-1}(t)q_{m-1} - p_{nm}(t)q_m \quad \text{for } m \geq n, \tag{22}$$

and

$$p'_{nm}(t) = q_n p_{n+1,m}(t) - q_n p_{nm}(t) \quad \text{for } m \geq n \tag{23}$$

with the convention that $q_0 = 0$ and the boundary condition

$$p_{nm}(0) = \begin{cases} 0 & \text{if } m \neq n, \\ 1 & \text{if } m = n. \end{cases}$$

It is easy to check that the following probabilities satisfy Equations (22) and (23):

$$p_{nm}(t) = \begin{cases} e^{-q_n t} & \text{if } m = n, \\ q_n \cdots q_{m-1} \displaystyle\sum_{k=n}^{m} \frac{e^{-q_k t}}{\displaystyle\prod_{j=n, j\neq k}^{m} (q_j - q_k)} & \text{if } m > n, \\ 0 & \text{if } m < n. \end{cases} \tag{24}$$

We follow the usual notation and let $P(t)$ denote the $S \times S$ matrix with entries $p_{nm}(t)$ especified in Equation (24).

It is worth pointing out that the transition probability $p_{nm}(t)$ with $m > n$ becomes undefined if $\prod_{j=n, j\neq k}^{m} (q_j - q_k) = 0$. Evidently, this occurs if there are at least two intensities q_j and q_k such that $q_j = q_k$.

Suppose the intensities of $\{X(t)\}_{t\geq 0}$ are given by $q_n = \frac{\lambda n(N-n)}{N-1}$ for all n in S. It is easily seen that $q_n = q_{N-n}$ holds for every state n. Therefore the transition probabilities $p_{nm}(t)$ are not defined for $m > \frac{N}{2}$ if N is even, or for $m > \frac{N-1}{2}$ if N is odd. This case occurs when every immigrant in the network has a search process of rate λ, and consequently the network's search process is a simple birth of rates $\lambda_n = n\lambda$.

To avoid singularities in the semigroup, we have refrained from this case by diminishing the rate of each immigrant through a fixed parameter δ. At the same time the insertion of this parameter follows technical reasons, we believe it makes the migration model more realistic. This is because the assumption of diminishing search rates illustrates that longer-term arrived immigrants have more information and more contacts in the host destination and this allows them to observe jobs with a higher frequency than newly added members. Conversely, when the search rates are equal, immigrants have the same average observation interarrival times. This result somewhat ignores what the migration literature says about the improvement of immigrant information systems as the elapsed time since migration increases (Todaro, 1969).

From Equation (24), we see that transition in $t = 0$ yields

$$P(0) = I, \quad \text{where } I \text{ is the } S \times S \text{ identity matrix.} \tag{25}$$

Note that Equation (25) together with the observation that (24) satisfies (22) actually proves that the family $\{P(t) : t \geq 0\}$ is the transition semigroup of the birth process $\{X(t) : t \geq 0\}$.

Appendix C Density and distribution functions of the time until absorption

We follow the notation used in Mogens (2005) throughout this Appendix.

Let π_n denote the probability that the counting process of immigrants $\{X(t)\}_{t\geq 0}$ starts in state n for all n in S. Given time is measured from the arrival of the first immigrant we have $\pi_1 = \mathbb{P}(X(0) = 1) = 1$ and $\pi_n = \mathbb{P}(X(0) = i) = 0$ for all $n > 1$. The initial distribution of $\{X(t)\}_{t\geq 0}$ (defined on its transient states) is then $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{N-1}) = (1, 0, \dots, 0)$.

The infinitesimal generator of $\{X(t)\}_{t\geq 0}$ is given by the matrix

$$Q = \begin{pmatrix} \boldsymbol{\Theta} & \boldsymbol{t} \\ \boldsymbol{0} & 0 \end{pmatrix}, \tag{26}$$

where $\boldsymbol{\Theta}$ is the subgenerator matrix

$$\boldsymbol{\Theta} = \begin{pmatrix} -q_1 & q_1 & 0 & \cdots & 0 & 0 \\ 0 & -q_2 & q_2 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & -q_{N-2} & q_{N-2} \\ 0 & 0 & \cdots & 0 & 0 & -q_{N-1} \end{pmatrix},$$

$\boldsymbol{t}$ is a $(N-1)$-dimensional column vector $\boldsymbol{t} = \begin{pmatrix} 0 & 0 & \cdots & 0 & q_{N-1} \end{pmatrix}^T$, and $\boldsymbol{0}$ a $(N-1)$-dimensional vector of zeros. Under the assumptions stated above, the time until absorption τ has a phase-type distribution with initial distribution $\boldsymbol{\pi} = (1, 0, \dots, 0)$ and subgenerator matrix $\boldsymbol{\Theta}$.[30]

For the general case, the density function of the time until absorption of a Markov jump process when $\tau \sim \text{PH}(\boldsymbol{\pi}, \boldsymbol{\Theta})$ is

$$f_\tau(x) = \boldsymbol{\pi} e^{\boldsymbol{\Theta} x} \boldsymbol{t}, \tag{27}$$

[30] See Mogens, 2005.

where $e^{\boldsymbol{\Theta} x}$ is the transition probability matrix to the transient states of $\{X(t)\}_{t\geq 0}$ at time-lag x .

Substituting the initial distribution $\boldsymbol{\pi} = (1, 0, \ldots, 0)$ and $\boldsymbol{t} = \begin{pmatrix} 0 & \cdots & 0 & q_{N-1} \end{pmatrix}^T$ into Equation (27) yields

$$f_\tau(x) = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} p_{11}(x) & p_{12}(x) & \cdots & p_{1,N-2}(x) & p_{1,N-1}(x) \\ 0 & p_{22}(x) & \cdots & p_{2,N-2}(x) & p_{2,N-1}(x) \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & 0 & p_{N-1,N-1}(x) \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ q_{N-1} \end{pmatrix}.$$

From this it follows that the density function of τ is given by

$$f_\tau(x) = p_{1,N-1}(x) q_{N-1}. \tag{28}$$

The value of $p_{1,N-1}(x)$ can be directly obtained from (24), combining it with Equation (28) we obtain

$$f_\tau(x) = \sum_{i=1}^{N-1} q_i e^{-x q_i} \left(\prod_{j=1, j\neq i}^{N-1} \frac{q_j}{q_j - q_i} \right).$$

This establishes that τ has hypoexponential distribution of parameters $q_1, \ldots, q_{N-1}$.

The task is now to find the distribution of function of τ. We begin with a general result on the distribution function of the time until absorption of a Markov process:

$$F_\tau(x) = 1 - \boldsymbol{\pi} e^{\boldsymbol{\Theta} x} \boldsymbol{e}. \tag{29}$$

Note that in our case,

$$\boldsymbol{\pi} e^{\boldsymbol{\Theta} x} \boldsymbol{e} = \sum_{i=1}^{N-1} p_{1i}(x).$$

As $P(x)$ is a stochastic matrix we have $\sum_{j=1}^{N} p_{ij}(x) = 1$ for all i in S. By virtue of this we can rewrite $\boldsymbol{\pi} e^{\boldsymbol{\Theta} x} \boldsymbol{e}$ as $1 - p_{1N}(x)$. Substituting $\boldsymbol{\pi} e^{\boldsymbol{\Theta} x} \boldsymbol{e}$ into (29) we see that

$$F_\tau(x) = p_{1N}(x). \tag{30}$$

On substituting the value of $p_{1N}(x)$ into (30) we obtain

$$F_\tau(x) = \prod_{i=1}^{N-1} q_i \sum_{i=1}^{N} \frac{e^{-q_i x}}{\prod_{j=1, j\neq i}^{N} (q_j - q_i)},$$

which provides an explicit description of the distribution function of τ.

Voting with preference for alternation

Latin American countries like Colombia, Guatemala, Mexico, and Paraguay have one term presidential limits in the hopes of preventing dictatorships and promoting partisan balance. The idea is that open seat races reduce incumbency advantages, and increase the opposition's chances of winning the presidency. In this regard, party alternation is thought to be both a defining aspect of democracy, and a mechanism for advancing democratization in countries with authoritarian regimes. More precisely, the prospect of alternation, and thus of losing office and power, makes parties create democratic institutions like government agencies, and electoral commissions, to protect them if and when, they find themselves on the opposition side (Maltz, 2007). An example of this occurred in Mexico, when the first opposition party that came to power in 2000 after dethroning the dominant party, proposed a law, which eventually lead to the creation of an institution, to make government information available to the public.

Party alternation is also linked to political accountability. As Carbone and Pellegata point out: "When elections lead to changes in government, the underlying implication is that the performance of the ousted government did not satisfy the voters' demands and expectations, and that the latter are thus requesting policy adjustments" (2017: 9).[1] Discontent towards the ruling party may arise from corruption or bribery scandals (Coşkun 2016), cronyism (Klor *et al.* 2017), or the lack of change in electoral reforms. Regarding this latter cause, Pilet and Bol (2011) analyze the effect of party tenure on the preference for political reform, and find that the longer parties are in power, the less likely they are to seek reforms.

It is not the case however, that term limits necessarily lead to party rotation. Take for example Mexico, who despite having a single six year term limit for presidents since 1934, kept the Institutional Revolutionary Party (PRI) in power for 66 consecutive years. Diaz-Cayeros and Magaloni (2001) argue that party dominance was due to electoral rules that disincentivized coordination among opposition parties, yet it is also true that part of the electorate had developed a sympathy, or at the very least, a tolerance for the ruling party. Voters often justified their choice for PRI candidates with the phrase "better the devil you know than the devil you don't". Hence, the almost 7 decade-long dominance of PRI could be partially explained by the voters' attitude towards risk; in particular, by a prevalence of individuals who shied away from uncertainty, and voted for PRI's official candidates, by a lack of individuals drawn to uncertainty who would have voted for challengers, or by a combination of both. The main reference is Eckles *et al.* (2013), who show that risk-averse individuals are prone to vote for incumbents, while risk accepting individuals are more likely to vote for challengers.

Support for the incumbent party's candidate can also be explained by partisan incumbency advantage, defined by Fowler and Hall as 'the electoral benefit a candidate receives purely because her party is the incumbent party, regardless of whether she herself previously served' (2014: 502). They distinguish two kinds of partisan advantage: positive, where of-

[1] While it would be tempting to link this to the research on constituency service, or federal money spending (see Gaines 1998, King 1991, and Klingensmith 2015), most studies focus on the effect of the incumbency of politicians, rather than that of parties. So, the even though balance of power is a relevant concept in this strand of literature, it is not the kind we wish to explore.

ficial candidates benefit from the voters' favor for incumbents, and negative, which signals that voters have a preference for partisan balance. Note that if positive partisan advantage is present at the time of elections, then political competition decreases, and alternation is not an immediate implication of open seat elections.

In this work, we introduce a concept closely related to risk aversion and partisan advantage called preference for alternation. Preference reflects the voters' evaluation of outside options, so it comprises the utility of voting for the opposition when a given party is incumbent. The evaluation may be driven by the voters' emotions, rational motivations, or a combination of both. Lerner *et al.* (2004) claim that emotions like anger or sadness can lead to risk-seeking choices, whereas disgust evokes avoidance to try something new. Emotions towards the incumbent party may arise from rational situations, like anger stemming from corruption scandals,[2] but also from exogenous factors of which politicians have no control. Wolfers (2002) shows that voters reward or punish governors for oil price raises, based on whether their state is an oil producer or consumer, in spite the fact that oil prices are thought to exhibit random behavior (Serletis and Andreadis, 2004).

We are interested in investigating to what extent preference for alternation affects the voters' evaluation of candidates and consequently, the results of elections. In addition, we analyze whether term limits can alleviate the confounding effect of preference for alternation. The idea is that uncertainty, and therefore heterogeneous attitudes of voters towards risk, arise from open seat elections, but re-elections carry a higher degree of information about at least one candidate: the incumbent. So, even if voters are prone to alternation, they may end up re-electing the incumbent if she proves to be competent. We emphasize that incumbents may not necessarily reveal their competence through their actions, see Aghion and Jackson (2016) for an account of the strategies available to principals to induce agents to reveal their types.

[2] An example would be Peru's former president Pedro Pablo Kuczynski, who was linked to vote-buying, and had to resign after videos of his allies bribing lawmakers were made public.

To the best of our knowledge, our study along that of Eckles *et al.* and Fowler, is one of the few that analyze incumbency from the perspective of voters, and not of politicians or political parties. Typically, the scholarly literature focuses on the mechanisms that lead to incumbency advantages such as redistricting and gerrymandering (Cox and Katz, 2002), campaign finance (Meirowitz, 2008), and returns to experience (Lee, 2001), or the way in which the advantage is perpetuated: either by positive electoral selection (Ashworth and Bueno de Mesquita, 2008), or by scaring off experienced challengers and lowering the quality of candidates from the opposition. Hall and Snyder (2015) find that this accounts only for 5-7 percent of the incumbency advantage in U.S. elections.

The closest reference to our work is that of Fowler (2018), where voters receive noisy signals about candidate quality, and update their beliefs about incumbents. Under uncertainty, incumbency is a positive signal that tends to favor incumbents, even when they are not intrinsically better than challengers. Just like Fowler, we analyze the effects of incumbency in terms of the voter's decision to elect an incumbent over a high quality challenger, but make a more detailed analysis of the implications of this for party rotation. Our dynamic setting also allows us to extend time beyond the finite period horizon, which is standard for games with voting, see Dekel and Piccione (2000), and Battaglini (2005) for a full treatment of sequential voting, and Jackson and Tan (2012) for information revelation via voting rules. This implies that we can analyze which party spends more time on average in power in the long-run.

In our model, voters elect the contender that they think is most suitable for office. The catch is that the voters' evaluation of candidates is affected by their own preference for political alternation. The characterization we make is that voters that prefer stability have an enhanced perception of the candidates of the ruling party because they want to keep it in power, while voters that have a preference for alternation think lower of them because they want to try something new. Candidates are completely characterized by valence, which is a variable that comprises individual conditions beyond the left-right Downsian ideological dimension, and that are positively valued by voters (Stokes, 1963). Roughly speaking, valence can be viewed as the intrinsic appeal of a candidate, irrespective of her own, or her party's policy positions. Some authors classify candidate valence issues into those that are useful in elections, and those useful for governing (Adams *et al.* 2011). We do not disaggregate valence into its components, but treat it as a general indicator of candidate appeal. On account of the fact that voters make their decision based on candidate appeal and not policy, in other words, because they are strictly *valence motivated*, our model cannot be situated in the classical Downsian coordinate system. Unlike other models that have introduced the half-dimension of valence into the voters' utility function (see Serra 2010 for a deeper discussion of the relationship between valence and policy), our model is situated in its own dimension.

Although the assumption that voters only value non-ideological characteristics of candidates may seem little realistic, empirical evidence suggests that a fair proportion of the electorate relies on factors different from economic or social policies when electing their leaders, and instead focuses on issues in other dimensions. As Stokes points out "A few voters, as we have seen, impose a clear ideological structure on political conflict. But the vast majority rely on assorted non-ideological ways of structuring the political world" (1963: 375). Confirming this view, is Nyhuis' (2016) finding that candidate valence,[3] is positively and significantly associated with vote share. Empirical studies have also shown that physical features of candidates are correlated with election outcomes. Todorov *et al.* (2005), and Ballew and Todorov (2007) show that the evaluation of candidate competence made by participants based on politicians' facial appearance, correctly predicted 68% of US gubernatorial races, and 72% of US Senate races. Voters that lack information about candidates tend to rely more on these judgments of quality based on physical attributes or characteristics. McDermott (1997) shows that in low-information elections, voters use gender as a cue for policy position, and in particular, finds that women candidates from the Democratic party are preferred over

[3]Nyhuis' indicator of candidate valence is the residual between vote recommendation based on the proximity between the preferred policy of voters and that campaigned by candidates, and actual vote.

men among liberal voters.

Even if voters are perfectly informed, and cast their vote in a rational way, there could be scenarios where valence is a tie-breaking determinant for choosing from similar alternatives. Downs (1957) suggests that as parties converge towards the same ideological point, there is less ground for distinction, and the voters' decision may boil down to 'some irrational basis', or in other words, a valence issue. Serra (2013) describes a scenario in the 2012 presidential election in Mexico, where the three main runners proposed a similar strategy to address security issues, namely, to keep the army in the streets (Mexico was experiencing considerable violence as a result of the war on drug traffickers declared by the former president). Although candidates differed in other concerns, this is an example where voters could have relied on criteria different from the candidates' policies to make their choice. We highlight the fact that campaign teams have also noticed the value of campaigning on valence, at least more than on ideology, if candidates lack certain attributes like charisma, or if the electorate is relatively sensitive to political ideology. Serra (2013, 2018) argues that Hillary Clinton campaigned on her experience in government, a *character valence* issue, to compensate for her lack of charisma, while Peña Nieto refrained from making explicit policy proposals, and instead relied on his appeal, which was also portrayed in different sources of media.

We assume that a voter against alternation, or prone to stability, thinks that the candidate of the ruling party has a higher valence than she actually does, while a voter in favor of alternation thinks that it is lower. A voter that is indifferent to alternation perceives valence without distortion. Note that preference for alternation has the same effect on voting as risk aversion or positive partisan advantage, while preference for stability is consistent with risk acceptance or negative partisan advantage. The difference is that partisan advantage is intrinsic to parties, even though it affects voter behavior, while risk aversion and preference for alternation pertain solely to voters.

In this setting, the lack of political alternation can be explained by a series of combinations of preferences for alternation and candidate valences. For instance, an uninterrupted winning streak, or party dominance, can occur if voters are indifferent to alternation and all the candidates of a given party have higher valences than their challengers. Alternatively, voters can be consistently against alternation, and their loyalty to the party in power can either boost strong official candidates, or compensate for less competitive ones. Another possibility is that voters are always drawn to alternation, but the candidates of the opposition lack sufficiently high valences to win office. Even though we have named a few examples in which the preference type of voters remains the same throughout time, we can consider scenarios in which the preference for alternation changes every election. In such dynamic settings, there are still combinations of preferences and valences that lead to consecutive victories of a single party.

We find that preference factors have a bigger effect on the probability of electing the lesser candidate in the two term limit, than they do when there is a single term limit. Explicitly, if the valence of an incumbent and that of a new candidate are affected by the same preference factor (different from the null), then it is more likely that voters elect the lowest valence valued candidate: this could be the challenger, if the incumbent has a higher valence, or vice versa. The reason being that when voters have a distorted evaluation of the incumbent, the information about her previous performance is discarded, and this involves a bigger loss of information than modifying the perception of a new candidate. This suggests that if voters have preferences for, or against alternation, then legislative single terms lead to higher average valence politicians in office. The closest empirical estimate of the preference of voters is partisan incumbency advantage, defined by Fowler (2014) as $\frac{W_j(1)-W_j(0)}{2}$, where $W_j(1)$ is the party vote share received by the Democratic party in an open seat election in district j where the previous incumbent was a Democrat, and $W_j(0)$ is the vote share received by the Democratic Party's candidate in the same district in the counter factual scenario where the incumbent was a Republican. Using U.S. state legislative elections data, he finds that partisan advantage is indistinguishable from zero, which in our case implies that preference for alternation is negligible. Regarding the average number of victories of parties, we find that two term limits decrease partisan balance, whereas single term limits carry more rotation, but the differences between the two schemes attenuate over time. This suggests that a policy on legislative terms can have a limited scope.

In Section 1 we present the election process with a single term limit and preference for alternation. We compute the probability of observing streaks, and analyze the long-run behavior of election results. In Section 2, we extend the model to allow for two consecutive terms in office. We include a detailed explanation of the computational challenges of the extension, and provide intuition for an analytical solution. We study the long-run behavior of elections through simulations. Section 3 deals with efficiency and provides an example of the effect of preference for alternation on the probability of choosing the less qualified candidate.

7 Single term model

The discrete-time model we consider is one where there are two political parties, A and B. Every period t in $\{0, 1, 2, \ldots\}$, a candidate from each party runs for office. It is assumed that candidates are completely characterized by valence, and that the electorate prefers high over low valence candidates. The valence of each candidate is a random variable that takes non-negative values. Let $\{\upsilon_t^i\}_{t\in\{0,1,\ldots\}}$ be the sequence of independent and identically distributed (i.i.d) valences of the candidates of party i in $\{A, B\}$. We emphasize that we do not require $\{\upsilon_t^A\}_{t\in\{0,1,\ldots\}}$ and $\{\upsilon_t^B\}_{t\in\{0,1,\ldots\}}$ to be drawn from the same distribution, this allows to model situations in which a party has higher valence candidates, or higher variance across valences. We make the assumption that υ_t^A is independent of υ_t^B for all $t \geq 0$.

Each period voters have a preference for party alternation, and this alters their evaluation of the incumbent party's candidate valence. Preference at time t, denoted by ε_t, is a random variable that takes strictly positive values, and affects the perception of the official candidate's valence via multiplication. So, if party i in $\{A, B\}$ is in power at time $t - 1$, voters value the valence of i's candidate at time t as $\varepsilon_t \upsilon_t^i$.[4] If the realization of ε_t is such that the valence evaluation $\varepsilon_t \upsilon_t^i$ is higher than υ_t^i, we regard preference as positive partisan incumbency advantage. If the evaluation is lower than the original valence, then preference acts as negative partisan advantage.

For simplicity, we assume that $\{\varepsilon_t\}_{t\in\{1,2,\ldots\}}$ are i.i.d random variables, and that ε_{t+1} is independent of υ_t^i for all $t \geq 0$. Both are strong assumptions, as the first implies that voters have no consistent attitudes towards incumbent parties, or in other words, there is no cumulative (dis)content effect from having the same party in office for a number of terms. The second assumption implies that the incumbent's valence and the voters' preference in the following period are uncorrelated. This implies that our model has no party reputation effect. In other words, if an incumbent performs exceptionally well, the following official candidate will not benefit from her good reputation, and likewise, if the incumbent performs badly, voters will not transfer the blame to the sitting party's nominee.

We make the assumption that any candidate that contends for office in a given period is not allowed to run in any subsequent period regardless of the outcome of the election. Timing is as follows.

1. At the beginning of period 0, the candidate with the highest valence wins the office.

[4] The implicit assumption is that voters are identical, and thus have the same preference for alternation. This simplifying condition can be relaxed by allowing each voter k in $\{1, 2, 3, \ldots, N\}$ to have her own preference factor ε_t^k. This extension is not particularly difficult to handle, as it implies there would be a finite number of valence evaluations $\{\varepsilon_t^1 \upsilon_t^i, \varepsilon_t^2 \upsilon_t^i, \ldots, \varepsilon_t^N \upsilon_t^i\}$ to be compared against the valence of the challenger, υ_t^j, with $j \neq i$ in $\{A, B\}$. More precisely, the candidate of party i in $\{A, B\}$ would win the election, if the number of voters that deem her more suitable for office than the challenger, namely $\sum_{k=1}^{N} \mathbb{1}\{\varepsilon_t^k \upsilon_t^i > \upsilon_t^j\}$, is bigger than the number of voters that prefer the challenger over her, $\sum_{k=1}^{N} \mathbb{1}\{\varepsilon_t^k \upsilon_t^i < \upsilon_t^j\} = N - \sum_{k=1}^{N} \mathbb{1}\{\varepsilon_t^k \upsilon_t^i > \upsilon_t^j\}$. We believe this result could be confounded by the size of the voting population N, and this is the reason why we have imposed the condition of voter homogeneity. Nevertheless, the analysis is possible, and, under proper handling of variables, could lead to interesting comparative statics with respect to the size of the electorate.

At the end of her term, the incumbent vacates her position.

2. In period $t \geq 1$, the valence of the incumbent party's candidate is drawn from the distribution of the product of the voter's preference and the ruling party's valence distribution. The valence of the challenger is drawn from her own party's valence distribution. So, if i in $\{A, B\}$ won the election in period $t-1$, the valence of the official candidate is $\varepsilon_t v_t^i$. The valence of the challenger, that is, the valence of j's candidate is v_t^j, with $i \neq j$. As before, the candidate with the highest valence wins the office. Once her term is over, the incumbent vacates her position.

Let x_t in $\{A, B\}$ be the ruling party of period t in $\{0, 1, 2, \ldots\}$. So, the vector $(x_0, x_1, \ldots, x_{t-1}, x_t)$ gives the sequence of winning parties from period 0 to period t. We denote by $X_t = x_t$ the t-th state of our process. It is worth pointing out that transitions from the current party in power to the next depend only on the present party in power, and not on the previous ones, thus $\{X_t\}_{t\in\{0,1,\ldots\}}$ is a Markov chain with state space $S = \{A, B\}$. The probability that the first party in power is x_0 is given by

$$\mathbb{P}(X_0 = x_0) = \begin{cases} \mathbb{P}(v_0^A \geq v_0^B) & \text{if } x_0 = A, \\ \mathbb{P}(v_0^B \geq v_0^A) & \text{if } x_0 = B. \end{cases}$$

It is of interest to know the effect of preference for alternation on political transitions. We first focus on winning streaks to determine under what conditions there may be an effect of party incumbency advantage. We later examine if preference for alternation affects the long-run number of victories of each party.

7.1 Streaks

Let us first note that by definition of the election process, the conditional transition probabilities are $\mathbb{P}(X_{t+1} = x_{t+1} \mid X_t = x_t) = \mathbb{P}(X_1 = x_1 \mid X_0 = x_0)$ for every $t \in \{0, 1, 2, \ldots\}$. As the transition probabilities are independent of t, and the valences $\{\upsilon_t^A\}_{t\in\{0,1,\dots\}}$ and $\{\upsilon_t^B\}_{t\in\{0,1,\dots\}}$ are identically distributed (respectively), the transition matrix of $\{X_t\}_{t\geq 0}$ denoted P, is

$$\boldsymbol{P} = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{bmatrix} \mathbb{P}(\varepsilon_1 \upsilon_1^A \geq \upsilon_1^B) & \mathbb{P}(\upsilon_1^B > \varepsilon_1 \upsilon_1^A) \\ \mathbb{P}(\upsilon_1^A > \varepsilon_1 \upsilon_1^B) & \mathbb{P}(\varepsilon_1 \upsilon_1^B \geq \upsilon_1^A) \end{bmatrix} \end{array}.$$

We emphasize that finding the distribution of the product of random variables, in this case $\varepsilon_1\upsilon_1^A$ and $\varepsilon_1\upsilon_1^B$, is by no means a trivial task. In general, if X and Y are two independent continuous random variables with respective probability density functions $f_X(x)$ and $f_Y(y)$, then the probability density function of $Z = XY$ is $f_Z(z) = \displaystyle\int_{-\infty}^{\infty} f_X(x) f_Y(z/x) \frac{1}{|x|} dx$. However, the implementation of this result is usually not straightforward.

Since our intention is to provide a general feel of voting with preference for alternation, we restrict our attention to easily tractable distributions of the product of random variables. In particular, we work with the easiest continuous product distribution we can think of: that of an exponential random variable and a positive constant. As the reader will see, even this simple case allows for multiple characterizations of candidate valence and preferences of voters. Moreover, since the family of exponential distributions is closed under scaling by a positive factor, the choice of exponential and constant variables allows for a different interpretation of the events leading to alternation. In what follows, preference for alternation ε_t is defined to be a constant random variable. That is, preference is a deterministic variable with degenerate cumulative distribution

$$F_{\varepsilon_t}(x) = \begin{cases} 0 & \text{if } x < c, \\ 1 & \text{if } x \geq c, \end{cases} \tag{31}$$

as mentioned before, this is assumed only for simplicity. However, we provide in Appendix G a program that samples preference factors from a customizable distribution for every period t, this allows us to obtain numerical results.

Proposition 4. *If the candidate valences υ_t^A and υ_t^B are independent exponential variables with respective parameters λ_A and λ_B, and voters prefer stability, so $\varepsilon_t > 1$, then the evaluation of the official candidate's valence $\varepsilon_t \upsilon$, is drawn from a higher average distribution than her original valence.*

The result asserts that voters prone to stability have on average, a higher evaluation of the incumbent party's candidate valence than she originally does.

Proof. Without loss of generality, let us condition on the event A being the ruling party at time $t-1 > 0$. Observe that by definition of the single term model, the voter's evaluation of A's candidate is $\varepsilon_t \upsilon_t^A$, while the valence of her challenger is simply υ_t^B. If $\varepsilon_t > 1$, then $\mathbb{E}(\varepsilon_t \upsilon_t^A) = \frac{\varepsilon_t}{\lambda_A} > \mathbb{E}(\upsilon_t^A) = \frac{1}{\lambda_A}$, which implies that valence evaluation of A's candidate is drawn from a higher average distribution than her original valence. Equivalently, the perceived valence of the challenger $\frac{\upsilon_t^j}{\varepsilon_t}$ is drawn from a lower average distribution than the valence υ_t^j. $\square$

The interest of Proposition 4 is in the assertion that a positive bias for the ruling party can be interpreted as a negative bias for the opposition. This implies that a single parameter like ε_t can potentially portray the voter's ideological preference, if such interpretation of the model were to be made.

Lemma 7.1. *If the valences of the candidates of A and B are exponential variables with respective rates λ_A and λ_B, then preference for stability increases the probability of observing uninterrupted winning streaks, while preferring alternation reduces it.*

Let us denote by S_n^A the uninterrupted winning streak of length $n+1$ that starts with party A in power at time 0, and that ends with party A in power at time n. We see that

$$\begin{aligned}
\mathbb{P}(S_n^A) &= \mathbb{P}(X_0 = A, X_1 = A, X_2 = A, \ldots, X_{n-1} = A, X_n = A) \\
&= \mathbb{P}(\upsilon_0^B \le \upsilon_0^A, \upsilon_1^B \le \varepsilon_1 \upsilon_1^A, \upsilon_2^B \le \varepsilon_2 \upsilon_2^A, \ldots, \upsilon_{n-1}^B \le \varepsilon_{n-1} \upsilon_{n-1}^A, \upsilon_n^B \le \varepsilon_n \upsilon_n^A) \\
&= \mathbb{P}(\upsilon_0^B \le \upsilon_0^A)\mathbb{P}(\upsilon_1^B \le \varepsilon_1 \upsilon_1^A)\mathbb{P}(\upsilon_2^B \le \varepsilon_2 \upsilon_2^A) \cdots \mathbb{P}(\upsilon_{n-1}^B \le \varepsilon_{n-1} \upsilon_{n-1}^A)\mathbb{P}(\upsilon_n^B \le \varepsilon_n \upsilon_n^A),
\end{aligned}$$

where the last equality is due to the fact that $\{\upsilon_t^A\}_{t\in\{0,1,2,\ldots\}}$, $\{\upsilon_t^B\}_{t\in\{0,1,2,\ldots\}}$, and $\{\varepsilon_t\}_{t\in\{1,2,3,\ldots\}}$ are independent variables.

Assuming that the valences are exponentially distributed with respective parameters λ_A and λ_B, and that $\{\varepsilon_t\}_{t\in\{1,2,3,\ldots\}}$ is a sequence of positive numbers, then $\mathbb{P}(S_n^A) = \frac{(\lambda_B)^{n+1}}{\prod_{i=0}^{n}\left(\frac{\lambda_A}{\varepsilon_i} + \lambda_B\right)}$, where $\varepsilon_0 \equiv 1$. If voters were indifferent to alternation, elections would become independent Bernoulli trials where the probability that A wins office is inversely proportional to the average valence of B's candidates.

7.2 Long run behavior

We wish to analyze the amount of time that parties spend in power as a function of the average candidate valence and the voters' preference for alternation. To do this, we compute the average number of victories of each party as time goes to infinity. Let π_i denote the long run proportion of victories of party i in $\{A, B\}$, so $\pi_i = \lim_{t\to\infty} \frac{\sum_{j=0}^{t} \mathbb{1}_{\{X_j=i\}}}{t}$.

Theorem 1. *If $\{\upsilon_t^A\}_{t\in\{0,1,2,\ldots\}}$ and $\{\upsilon_t^B\}_{t\in\{0,1,2,\ldots\}}$ are i.i.d exponential variables with respective parameters λ_A, λ_B, and $\varepsilon_t \equiv \varepsilon > 0$ for all t in $\{1, 2, 3, \ldots\}$, then the stationary distribution of $\{X_t\}_{t\in\{0,1,2,\ldots\}}$ is*

$$\pi = (\pi_A, \pi_B) = \left(\frac{\lambda_A\lambda_B + \varepsilon\lambda_B^2}{2\lambda_A\lambda_B + \varepsilon(\lambda_A^2 + \lambda_B^2)}, \frac{\lambda_A\lambda_B + \varepsilon\lambda_A^2}{2\lambda_A\lambda_B + \varepsilon(\lambda_A^2 + \lambda_B^2)}\right). \tag{32}$$

Proof. This follows from the fact that $p_A + p_B = 1$, $0 \le \pi_i \le 1$, and $\pi_i = \sum_{k\in\{A,B\}} \pi_k p_{ki}$, where the conditional transition probabilities p_{ki} are given by the transition matrix of the process $\{X_t\}_{t\in\{0,1,2,\ldots\}}$,

$$P = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \left[\begin{array}{cc} \frac{\varepsilon\lambda_B}{\lambda_A+\varepsilon\lambda_B} & \frac{\lambda_A}{\lambda_A+\varepsilon\lambda_B} \\ \frac{\lambda_B}{\lambda_B+\varepsilon\lambda_A} & \frac{\varepsilon\lambda_A}{\lambda_B+\varepsilon\lambda_A} \end{array}\right] \end{array}.$$

□

We mention briefly another interpretation of the stationary distribution. We have that $\pi_i = \frac{1}{\mu_i}$, where μ_i is the expected first return time to state i. The formula says that the long run proportion of victories of party i is inversely related to the expected number of elections before the next victory of i, given that its candidate was elected in period 0. The intuition is that the faster it takes to i to return to power, the higher its average number of victories in the long run.

An interesting interpretation of μ_i is that it gives, by definition, the average length of j's first winning streak. Indeed, the events before i's first return to power correspond to uninterrupted victories of j. While this is true for the first streak, it extension to the rest of the streaks is immediate, as transitions depend only on the states, and not on time (this is due to the assumption of constant preference factors). This means that $\mu_A = \frac{2\lambda_A\lambda_B+\varepsilon(\lambda_A^2+\lambda_B^2)}{\lambda_A\lambda_B+\varepsilon\lambda_B^2}$ gives the expected length of B's winning streak, while $\mu_B = \frac{2\lambda_A\lambda_B+\varepsilon(\lambda_A^2+\lambda_B^2)}{\lambda_A\lambda_B+\varepsilon\lambda_A^2}$ gives that of A.

From now on, we continue with the interpretation of the stationary distribution as the long-run average number of victories of parties, and refer the interested reader to Norris (1998) for more details of the Ergodic theorem for Markov chains.

7.2.1 Analysis

From (32), we have that $\frac{\partial \pi_i}{\partial \lambda_i} < 0$, this means that when the average valence of the candidates of a political party increases, so does its long-run time in office. We also have that $\frac{\partial \pi_i}{\lambda_j} > 0$, which implies that when the average candidate valence of a party increases, the proportion of victories of the opposition decreases.

Perhaps less evident is the effect of the voters' preference on the amount of time that parties spend in power, given the non-monotonic behavior of π_i with respect to ε. We proceed to analyze the preference value that maximizes a A's time in power given its valence parameter and that of B. This involves no loss of generality, as the stationary entries are symmetric. There are three cases to consider.

Corollary 7.1. *If* $\lambda_A = \lambda_B$, *then* $\frac{\partial \pi_A}{\partial \varepsilon} = 0$.

The corollary asserts that if both parties have the same average valence, then the voters' preference for alternation has no significant effect in the average tenure of each party in government. The intuition is the following. First note that ceteris paribus, if the voters' preference for stability ε grows, the probability that the official candidate wins her election $\mathbb{P}(X_{t+1} = A \mid X_t = A) = \mathbb{P}(X_{t+1} = B \mid X_t = B) = \frac{\varepsilon}{1+\varepsilon}$ increases, and the probability that the challenger wins $\mathbb{P}(X_{t+1} = B \mid X_t = A) = \mathbb{P}(X_{t+1} = A \mid X_t = B) = \frac{1}{1+\varepsilon}$ decreases.

Accordingly, we expect to see long uninterrupted winning streaks for large values of ε. Nevertheless, since the probability of political alternation is positive $\frac{1}{1+\varepsilon} > 0$ for all values of ε, we expect to see an eventual transition of power to the opposition. Once this transition occurs, the voters' preference for stability will keep the party in power for a long time until another transition is made.

Conversely, if voters have a strong preference for partisan balance, that is if ε is small, we expect to have short winning streaks and frequent alternations between A and B. Not surprisingly, the steady state distribution of the chain is $\pi = (\frac{1}{2}, \frac{1}{2})$ for all $\varepsilon > 0$. This is nothing but the statement that regardless of the preference for alternation, each party spends on average half of the time in power. Figure 3 illustrates the symmetry of the election chain with equal valence distributions.

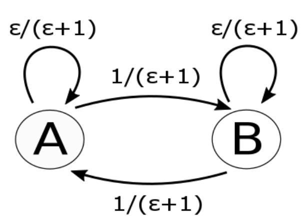

Figure 3: Equal valence chain

Corollary 7.2. *If* $\lambda_A < \lambda_B$, *then* $\frac{\partial \pi_A}{\partial \varepsilon} > 0$.

This result states that the party with the highest average valence benefits most from preference for stability. In particular, note that when the preference factor is very large, the long run proportion of victories of A becomes $\lim_{\varepsilon \to \infty} \pi_A = \dfrac{\lambda_B^2}{\lambda_A^2 + \lambda_B^2}$. This proportion is maximized when λ_A is infinitesimally small, that is, when A's average candidate valence extremely large. Indeed, in the limit when $\lambda_A \to 0$ and $\varepsilon \to \infty$, we have $\pi_A = 1$. In this case, we can think of party A as being an absorbing state of the election chain, in the sense that once a candidate of A is elected, its extremely well qualified official candidates will be elected with very high probability.

Therefore, the long run proportion of victories of each party, $\pi = (\pi_A, \pi_B)$, depends on the result of the first election. The stationary distribution of the chain is not unique, as all the vectors of the form $\pi = (\alpha, 1 - \alpha)$ with $0 \leq \alpha \leq 1$ satisfy $\pi \boldsymbol{P} = \pi$. Intuitively, we can think of this as a Bernoulli distribution stating the probabilities of flipping a coin and choosing the winning party in period 0.

Figure 4 presents the diagram of the chain with a low preference for alternation.

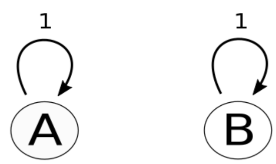

Figure 4: High valence and low preference for alternation chain

One sees immediately that each state communicates only with itself, that is, in the limit when voters are very much against alternation, there is a single dominant party.

Corollary 7.3. *If* $\lambda_A > \lambda_B$*, then* $\frac{\partial \pi_A}{\partial \varepsilon} < 0$.

If party B has higher valence-valued candidates than party A, the latter wants voters to have a strong preference for alternation. In particular, A benefits most from a scenario where voters alternate every period. This preference may seem counter intuitive, as constant alternation prevents long winning streaks. More precisely, a strong preference for alternation virtually eliminates the possibility that parties make self-transitions, and instead places this weight on transitions to the opposition. Nevertheless, A knows that its relatively weak contenders have little chance of successively beating B's challengers (let alone if these latter enjoy the preference of voters, that is if $\varepsilon > 1$). Thus, A prefers to have constant rotation (in other words, it prefers voters to have a preference such that $\varepsilon < 1$) as opposed to long losing streaks.

Note that we can interpret a 'negative' preference factor $\varepsilon < 1$ as discontent towards the ruling party, since both have the effect of diminishing the appeal of official candidates.[5] In this particular scenario, discontent may arise from voters having extremely high expectations for the incoming party. Serra (2013) argues for instance, that Vicente Fox, the first president from the opposition, faced unrealistic expectations of the electorate, and that despite many political and economics improvements, could not fully deliver on them.

Given the periodic behavior of this process, it is not surprising that A wins half of the elections and B the other half. The stationary distribution of this limiting case is given by $\pi = (\frac{1}{2}, \frac{1}{2})$.

The periodicity of this election chain is represented in Figure 5.

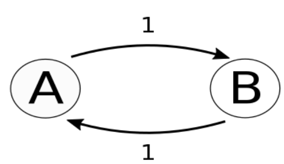

Figure 5: Low valence and high preference for alternation chain

[5]See Serra (2018) for an account of a popular discontent parameter δ in $[0, 1]$ that affects candidate charisma. The main difference is that his parameter affects the valence of challengers, while ours influences the valence of official candidates.

8 Two term model

We continue to make the assumptions of the no re-election model, and assume that incumbents are nominated for re-election. To simplify notation, we assume that whenever an incumbent runs for re-election, the candidate that would have run had the incumbent not been nominated is lost. So, if B's candidate is elected in period 0, then she runs for re-election in period 1, and the candidate whose valence is υ_1^B is forgotten. Timing is as follows.

1. At the beginning of period 0, the candidate with the highest valence is elected to office.

2. The incumbent is nominated to the election. Her valence evaluation is the product of her original valence, and the voter's preference factor. Her challenger is a new candidate from the opposing party. The nominee with the highest valence wins the election.

3. There are two cases to consider in period $t \geq 2$.

   3.1. The incumbent was re-elected in period $t - 1$. Given that the incumbent is ineligible to seek re-election to a third term, the winning party selects a new nominee at the beginning of period t. Her valence is multiplied by the preference factor ε_t. Her challenger is also chosen at the beginning of period t. The nominee with the highest valence is elected. The process continues as described in 2.

   To illustrate the dynamics, suppose that B's candidate is elected in period 0. In period 1 she runs for re-election, her perceived valence is $\varepsilon_1 \upsilon_0^B$, and that of her challenger is υ_1^A. In the event that she gets elected, she remains in office for another term and in the next period, her party picks a new candidate. Her valence evaluation is $\varepsilon_2 \upsilon_2^B$.

   3.2. The challenger won the election in period $t - 1$. In this case, the incumbent is nominated by its party to the election in period t. The process continues as described in 2.

It is worth noting that the variable X_t, defined to be the ruling party of period t in $\{0, 1, 2, \dots\}$ does not comprise all the information needed to predict future election outcomes. To name a simple example, suppose that A's candidate is incumbent in period 1, so $X_1 = A$. This can happen in two ways: either A's candidate was elected in period 0 and then re-elected, or B's candidate was elected first, and lost re-election to A. Both cases imply that the valence of A's candidate will be affected by the preference factor ε_2 in period 2, but there is an important difference: for the next election, the latter case conveys more information about the incumbent than the former.

To be precise, the path $X_0 = A \to X_1 = A$ gives information about the incumbent's competitiveness, but this information will be lost because the incumbent is in her second term, which means that a new candidate will run in period 2. In contrast, the path $X_0 = B \to X_1 = A$ implies that the valence of the incumbent was greater than the distorted valence of B's successor, and this is relevant since that the incumbent will run for re-election in the following period. In this example, one sees that $X_1 = A$ alone does not allow to unequivocally determine the conditions under which the next election will take place, this is nothing but the statement that $\{X_t\}_{t\in\{0,1,\dots\}}$ does not satisfy the Markov property. Simply put, the Markov property says that given the present information, the information of the past is irrelevant to know what will happen in the future.

Incidentally, the example sheds light on a crucial aspect of the re-election dynamics: every election following the second term of an incumbent has two new candidates from each party, and their valences are independent of those of previous contenders. This means that the only relevant information about the ruling party sequence, no matter how long it may be, is the party affiliation of the last office holder. In other words, the process renews itself every period that follows a re-election. We present here three trajectories of different lengths, all of which involve the re-election of B's incumbent at some given point in time. The vertical lines indicate that there is a renewal, or that we can dispense with the sequence to the left, as long as we keep in mind that in the following period, voters have a distorted perception of B's candidate valence.

$$BB \mid AB \cdots \qquad ABABB \mid BA \cdots \qquad ABA \underbrace{\cdots}_{k} ABB \mid AA \cdots$$

We emphasize that the sequences AA and BB unambiguously indicate the re-election of the incumbent only in the case where A or B were elected in the first period. To see why this may not always be a good indicator, consider the simple case where B's candidate is elected the first three periods, so $X_0 = B, X_1 = B, X_2 = B$. If we were to make a note of the party in power at time 2 and go backwards in time to check if the same party was in power the previous period we would be tempted to mark a renewal in period 2, $BBB \mid$, whereas an extra step behind would have shown us that the renewal occurred in the previous period $BB \mid B$.

8.1 Long-run behavior

The renewal property of re-election allows us to condense history and to deal with a smaller state space than we would have had initially. What remains to show is that re-elections occur in a finite time, in which case, we can guarantee renewals.

Theorem 2. *If $\varepsilon_t \geq 1$ for all $t \geq 1$, the probability of observing an infinite cycle of perfect alternation is zero.*

Proof. We refer the reader to Appendix A. □

The theorem asserts that under the condition described above, there will be at least one re-election in the long-run. This is due to the fact that preference for stability increases the appeal of incumbents. If voters are indifferent to alternation, the probability of perfect alternation decreases with every passing period. The idea is that at time t, there are 2^t party sequences, but only two of them guarantee perfect alternation, that is $ABAB\cdots BA$ and $BABA\cdots AB$. As time elapses, the total number of outcomes increases geometrically, but the number of favorable events remains the same. This leads to the result that, as t tends to infinity, the probability of perfect alternation becomes negligible. We emphasize that even though Theorem 3 assures a re-election will occur, it does not specify when we might expect it.

8.1.1 Analysis

It would be desirable to obtain analytical formulas for the convergence of the two term process, but due to the complexity of the state space, we have not been able to do this. Nevertheless, we present a few simulations of elections that lead to some numerical results. The program, which is reproducible in R is provided in Appendix G. We compare these results with the predictions of the single term model, and discuss the differences and similarities between them.

In these simulations, we assume that both parties draw the valence of their candidates from an exponential distribution with rate 1. Partisan advantage is drawn from 4 uniform distributions with respective parameters $\left(0, \frac{1}{2}\right)$, $(0, 1)$, $(0, 2)$, and $(0, 4)$. For each advantage distribution, we run 100 simulations of elections with re-election dynamics, and compute the average number of victories of party A. Figure 6 plots the per period averages. The more continuous lines correspond to low advantage parameters, while the more spaced lines correspond to higher parameters.

We see that the continuous lines have a faster convergence to 0.5 than the spaced ones do, nevertheless convergence is apparent. We deduce that the higher the preference for alternation, the fastest the balance of power is achieved, while the higher the preference for stability, the longer it takes to reach a point of equity.

Recall that the result on single term with equal valence is that each party wins half of the elections in the long run, independent of the partisan advantage. This is consistent with our

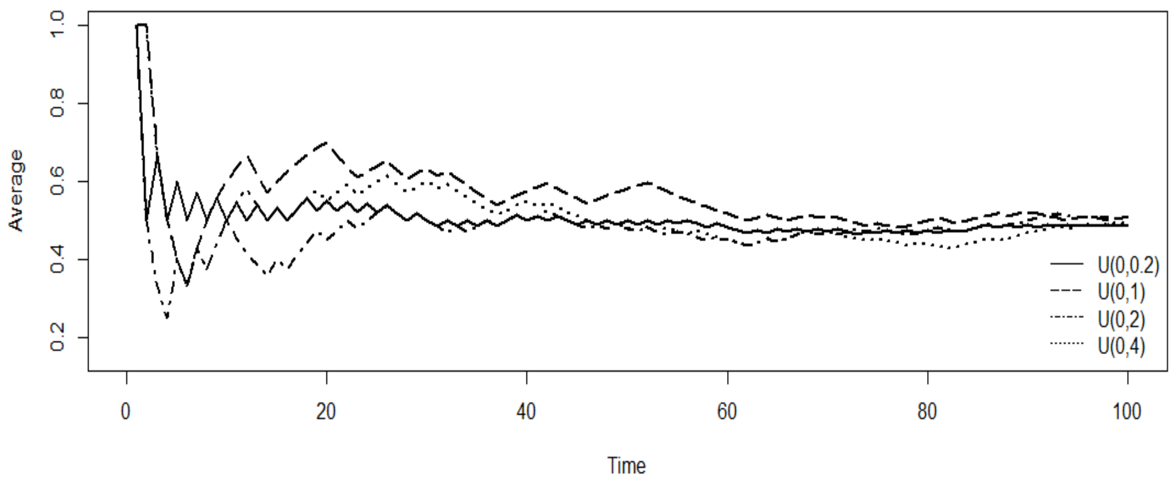

Figure 6: Stationary state with equal valence distributions

findings for the two term, where all the graphs end up reaching the same stationary state.

A straightforward comparison of the term schemes is presented in figure 7, where we plot the average number of victories of A for the single and two term dynamics. We assume that the average candidate valence of both parties is 1, and that the voters' preferences for alternation are uniformly distributed between 0 and 2.

One sees that the single term scheme reaches the stationary state 0.5 faster than the two term limit does. Despite this fact, the behavior of both functions is remarkably similar. We attribute the difference in the averages to electoral selection, but note that as time elapses, the difference reduces significantly. A possible explanation is that voters do not account for the margin of victory, and this may create an incumbency effect similar to that in Fowler (2018), where Bayesian updating about the incumbent beliefs leads to a higher chance of electing the incumbent for a second term, even if the challenger is equally, if not better suited for office.

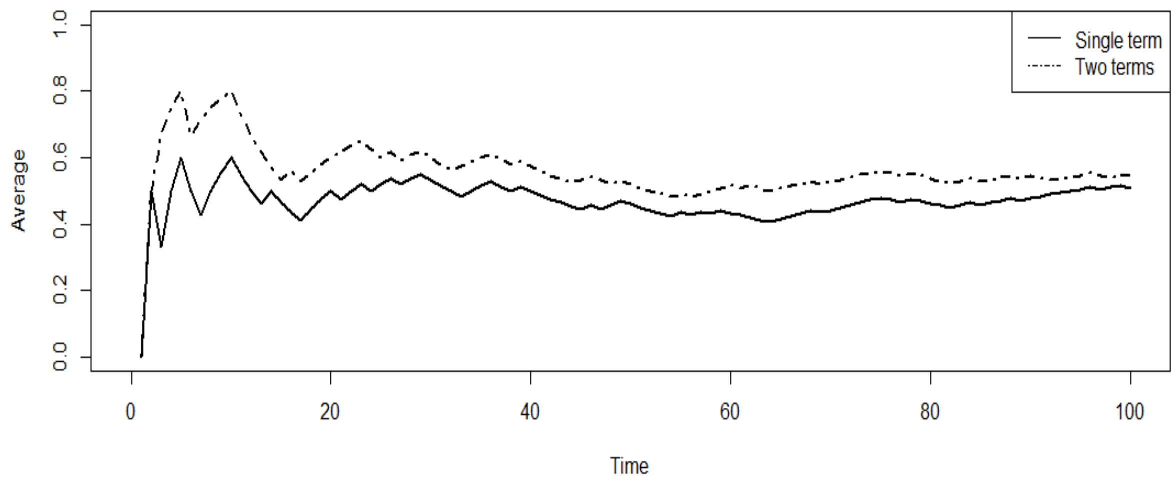

Figure 7: Stationary state with equal valence distributions

9 Efficiency

We wish to investigate if any of the term limit cases has a lower probability of electing the candidate with the lowest valence. Given that the calculations for the two term model are quite involved, we present an illustrative example of the election of the less qualified candidate in the second period. We do not do a formal generalization but present intuitive arguments for it. In what follows, we make the assumptions that the valences of A and B's candidates are exponential variables with respective parameters λ_A and λ_B, we also assume that the voters' preference in period 1 is the positive constant ε.

There is no loss of generality in assuming that the candidate of party B is elected in the first period. We are interested the probability that in the next election, voters choose the candidate with the lowest valence. There are two cases to consider:

1. Voters prefer alternation, $\varepsilon < 1$.

   In this case, voters have an enhanced perception of the challenger's valence in period 1.

An inefficiency would arise if they were to elect the challenger over the official candidate, whether it be the incumbent herself or a new candidate, if the official candidate has in fact a higher valence than the challenger. In other words, if voters choose what is new just for the sake of it being different, but not better. We present the probability of the aforementioned event for each term limit case.

(a) Single term

$$\mathbb{P}(X_1 = A \mid X_0 = B, v_1^A < v_1^B) = \frac{(1-\varepsilon)\lambda_B}{\varepsilon\lambda_A + \lambda_B}. \tag{33}$$

(b) Two terms

$$\mathbb{P}(X_1 = A \mid X_0 = B, v_1^A < v_0^B) = \frac{(1-\varepsilon)\lambda_B}{\varepsilon\lambda_A + \lambda_B} \cdot \frac{\varepsilon\lambda_A + 2(\lambda_A + \lambda_B)}{\varepsilon\lambda_A + (\lambda_A + \lambda_B)}. \tag{34}$$

Both probabilities have negative ε-derivative, which means that the probability of electing the less qualified contender is decreasing in the preference for alternation. The reason for this is that as ε tends to one, the bias towards the challenger decreases, so voters have a clearer perception of her true valence. The result is (33)<(34) for all values of λ_A, λ_B, and $\varepsilon < 1$.

2. Voters prefer stability, $\varepsilon > 1$.

In this case, voters have an enhanced perception of the incumbent party's candidate in period 1. An inefficient thing would be to elect her over the challenger if the challenger is better suited for office. We present the probability of this event for each term limit case.

(a) Single term

$$\mathbb{P}(X_1 = B \mid X_0 = B, v_1^B < v_1^A) = \frac{(\varepsilon-1)\lambda_A}{\varepsilon\lambda_A + \lambda_B}. \tag{35}$$

(b) Two terms

$$\mathbb{P}(X_1 = B \mid X_0 = B, v_0^B < v_1^A) = \frac{(\varepsilon-1)\lambda_A}{\varepsilon\lambda_A + \lambda_B} \cdot \frac{\varepsilon\lambda_A + 2(\lambda_A + \lambda_B)}{\varepsilon\lambda_A + \lambda_A + \lambda_B}. \tag{36}$$

Both probabilities have positive ε-derivative, which means that the probability of electing the lesser candidate is increasing in the preference for stability. In fact, as the ε factor grows, the bias towards the official candidate increases and the less realistic the perception of her valence becomes (on average). We have that (36)>(35) for all values of λ_A, λ_B and $\varepsilon > 1$.

We conclude from these examples that the preference factor ε has a stronger effect on the probability of electing the candidate with the lowest valence in the two term case than it does in the single term limit. The intuition is that in this scheme, voters observe the true valence of the incumbent in period 0. Any preference factor in the following period different from that of indifference ($\varepsilon = 1$) has the effect of distorting the acquired information, which, as we have mentioned is the true one. Simply put, knowing the truth and then modifying it has more consequences than not knowing the truth about something new and taking a guess.

References

[1] Adams, James, Samuel Merrill III, Elizabeth N. Simas, and Walter J. Stone. 2011. When candidates value good character: A spatial model with applications to congressional elections. *The Journal of Politics* 73, no. 1: 17-30.

[2] Aghion, Philippe, and Matthew Jackson. 2016. Inducing leaders to take risky decisions: dismissal, tenure, and term limits. *American Economic Journal: Microeconomics* 8, no. 3: 1-38.

[3] Ashworth, Scott, and Ethan Bueno de Mesquita. 2008. Electoral selection, strategic challenger entry, and the incumbency advantage. *The Journal of Politics* 70, no. 4: 1006-1025.

[4] Ballew, Charles C. II, and Alexander Todorov. 2007. Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 46: 17948-17953.

[5] Battaglini, Marco. 2005. Sequential voting with abstention. *Games and Economic Behavior.* 51: 445-463.

[6] Carbone, Giovanni, and Alessandro Pellegata. 2017. To Elect or Not to Elect: Leaders, Alternation in Power and Social Welfare in Sub-Saharan Africa. *The Journal of Development Studies* 53, no. 12.

[7] Coşkun, Gülçin Balamir. 2016. Does Corruption End the Dominant Party System? A Comparative Analysis of the Italian and Turkish Cases. *European Review* 25, no. 2: 320-336.

[8] Cox, Gary W., and Jonathan N. Katz. 2002. *Elbridge Gerry's Salamander: the electoral consequences of the reappointment revolution.* Cambridge University Press.

[9] Dekel Eddie, and Michele Piccione. 2000. Sequential voting procedures in symmetric binary elections. *Journal of Political Economy* 108, no. 1: 34-55.

[10] Diaz-Cayeros, Alberto, and Beatriz Magaloni. 2001. Party Dominance and the Logic of Electoral Design in Mexico's Transition to Democracy. *Journal of Theoretical Politics* 13, no. 3: 271-293.

[11] Downs, Anthony. 1957. An Economic Theory of Political Action in a Democracy. *Journal of Political Economy* 65: 135-150.

[12] Eckles, David L., Cindy D. Kam, Cherie L. Maestas, and Brian F. Schaffner. Risk attitudes and the incumbency advantage. Political Behavior 36: 731-749.

[13] Fowler, Anthony. 2014. Disentangling the personal and partisan incumbency advantages: evidence from close elections and term limits. *Quarterly Journal of Political Science* 9: 501-531.

[14] Fowler, Anthony. 2018. A Bayesian explanation for the effect of incumbency. *Electoral studies* 53: 66-78.

[15] Gaines, Brian J. 1998. The Impersonal Vote? Constituency Service and Incumbency Advantage in British Elections, 1950-92. *Legislative Studies Quarterly* 23, no. 2: 167-195.

[16] Hall, Andrew B., and James N. Snyder Jr. 2015. How much of the incumbency advantage is due to scare-off? *Political Science Research and Methods* 3, no. 3: 493-514.

[17] Jackson, Matthew and Xu Tan. 2012. Deliberation, disclosure of information, and voting. *Journal of Economic Theory* 148: 2-30.

[18] King, Gary. 1991. Constituency Service and Incumbency Advantage. *British Journal of Political Science* 21, no. 1: 119-128.

[19] Klingensmith, James Zachary. 2018. Using tax dollars for re-election: the impact of pork-barrel spending on electoral success. *Constitutional Political Economy.*

[20] Klor, Esteban F., Sebastian M. Saiegh, and Shanker Satyanath. 2017. Cronyism in State Violence: Evidence from Labor Repression During Argentina's Last Dictatorship. Available at SSRN: https://ssrn.com/abstract=2851822

[21] Lee, David S. 2001. The Electoral Advantage to Incumbency and Voters' Valuation of Politicians' Experience: A Regression Discontinuity Analysis of Elections to the U.S. House. August. NBER Working Paper 8441.

[22] Lerner, Jennifer S., D. A. Small, and G. Loewenstein. 2004. Heart strings and purse strings: carryover effects of emotions on economic decisions. *Psychological Science* 15, no. 5: 337-341.

[23] Maltz, Gideon. 2007. The Case for Presidential Term Limits. *Journal of Democracy* 18, no. 1: 128-142.

[24] McDermott, Monika L. 1997. Voting Cues in Low-Information Elections: Candidate Gender as a Social Information Variable in Contemporary United States Elections. *American Journal of Political Science* 41, no. 1: 270-283.

[25] Meirowitz, Adam. 2008. Electoral Contests, Incumbency Advantages, and Campaign Finance. *The Journal of Politics* 70, no. 3: 681-699.

[26] Norris, James R. 1998. *Markov chains.* Cambridge University Press.

[27] Nyhuis Dominic. 2016. Electoral effects of candidate valence. *Electoral Studies* 42: 33-41.

[28] Pilet, Jean-Benoit, and Damien Bol. 2011. Party Preferences and Electoral Reform: How Time in Government Affects the Likelihood of Supporting Electoral Change. *West European Politics* 34, no. 3: 568-586.

[29] Serletis, Apostolos, and Ioannis Andreadis. 2004. Random fractal structures in North American energy markets. *Energy Economics* 26, no. 3: 389-399.

[30] Serra, Gilles. 2010. Polarization of What? A Model of Elections with Endogenous Valence. *The Journal of Politics* 72, no. 2: 426-437.

[31] Serra, Gilles. 2013. Demise and Resurrection of a Dominant Party: Understanding the PRI's Comeback in Mexico. *Journal of Politics in Latin America* 5, no. 3: 133-154.

[32] Serra, Gilles. 2018. The electoral strategies of a populist candidate: Does charisma discourage experience and encourage extremism? *Journal of Theoretical Politics* 30, no. 1: 45-73.

[33] Stokes, Donald E. 1963. Spatial Models of Party Competition. *The American Political Science Review* 57, no. 2: 368-377.

[34] Todorov, Alexander, Anesu N. Mandisodza, Amir Goren, and Crystal C. Hall. 2005. Inferences of Competence from Faces Predict Election Outcomes. *Science* 308, issue 5728: 1623-1626.

[35] Wolfers, Justin. 2002. Are Voters Rational? Evidence from Gubernatorial Elections. *Stanford GSB Working Paper* no. 1730.

Appendix D

Theorem 3. *If $\varepsilon_t \geq 1$ for all $t \geq 1$, the probability of observing an infinite cycle of perfect alternation is zero.*

Proof. The idea of the proof is to compute the probability of observing a cycle of perfect alternation of any given length, and to show that it goes to zero in the limit when length goes to infinity. We begin by proving the following useful lemma.

Lemma D.1. *Let $X_1, X_2, \ldots, X_t$, be $t > 1$ i.i.d exponential variables with respective parameters $\lambda_1, \lambda_2, \ldots, \lambda_t$. Then,*

$$\mathbb{P}\left(X_1 \leq X_2 \leq \cdots \leq X_t\right) = \frac{\displaystyle\prod_{i=1}^{t-1} \lambda_i}{\displaystyle\prod_{i=1}^{t-1} \sum_{j=i}^{t} \lambda_j}$$

Proof. The proof is by induction on t. Consider $t = 2$, we have that $\mathbb{P}(X_1 \leq X_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$, so the statement holds for the first case.

Assume Lemma D.1 holds for k, so $\mathbb{P}(X_1 \leq X_2 \leq \cdots \leq X_k) = \dfrac{\prod_{i=1}^{k-1} \lambda_i}{\prod_{i=1}^{k-1} \sum_{j=i}^{k} \lambda_j}$, we prove it for $k+1 \leq t$. By definition,

$$\begin{aligned}
\mathbb{P}(X_1 \leq X_2 \leq \cdots \leq X_{k+1}) &= \int_0^\infty \int_{x_1}^\infty \cdots \int_{x_k}^\infty \lambda_1 \lambda_2 \cdots \lambda_{k+1} e^{-(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_{k+1} x_{k+1})} dx_{k+1} \cdots dx_2 dx_1 \\
&= \int_0^\infty \cdots \int_{x_{k-1}}^\infty \lambda_1 \cdots \lambda_k e^{-(\lambda_1 x_1 + \cdots + \lambda_k x_k)} \int_{x_k}^\infty \lambda_{k+1} e^{-\lambda_{k+1} x_{k+1}} dx_{k+1} \cdots dx_2 dx_1 \\
&= \int_0^\infty \cdots \int_{x_{k-1}}^\infty \lambda_1 \cdots \lambda_k e^{-(\lambda_1 x_1 + \cdots + \lambda_{k-1} x_{k-1} + (\lambda_k + \lambda_{k+1}) x_k)} dx_k \cdots dx_1 dx_1 \\
&= \frac{\lambda_k}{\lambda_k + \lambda_{k+1}} \left[\int_0^\infty \cdots \int_{x_{k-1}}^\infty \lambda_1 \cdots \lambda_{k-1} \lambda_{k'} e^{-(\lambda_1 x_1 + \cdots + \lambda_{k-1} x_{k-1} + \lambda_{k'} x_k)} dx_k \cdots dx_1 \right]
\end{aligned}$$

note that the term in brackets corresponds to $\mathbb{P}(X_1 \leq X_2 \leq \cdots \leq X_{k'})$, where $X_{k'}$ is an exponential random variable with parameter $\lambda_{k'} = \lambda_k + \lambda_{k+1}$. The induction hypothesis leads to

$$\mathbb{P}(X_1 \leq X_2 \leq \cdots \leq X_{k+1}) = \frac{\prod_{i=1}^{k} \lambda_i}{(\lambda_k + \lambda_{k+1})(\lambda_{k-1} + \lambda_k + \lambda_{k+1}) \cdots (\lambda_1 + \cdots + \lambda_{k+1})},$$

which is the desired conclusion. $\square$

Without loss of generality, we assume t is even. Let us denote by R_t^A, where R stands for rotation, the probability of observing a cycle of perfect alternation of length $t+1$ starting in state A. By lemma D.1, we have

$$
\begin{aligned}
R_t^A &:= \mathbb{P}(X_0 = A, X_1 = B, \ldots, X_{t-1} = B, X_t = A) \\
&= \mathbb{P}(\upsilon_0^B \leq \upsilon_0^A, \varepsilon_1 \upsilon_0^A \leq \upsilon_1^B, \ldots, \varepsilon_{t-1}\upsilon_{t-2}^A \leq \upsilon_{t-1}^B, \varepsilon_t \upsilon_{t-1}^B \leq \upsilon_t^A) \\
&= \mathbb{P}\left(\upsilon_0^B \leq \upsilon_0^A \leq \frac{\upsilon_1^B}{\varepsilon_1} \leq \frac{\upsilon_2^A}{\varepsilon_1\varepsilon_2} \leq \cdots \leq \frac{\upsilon_{t-1}^B}{\prod_{i=1}^{t-1}\varepsilon_i} \leq \frac{\upsilon_t^A}{\prod_{i=1}^{t}\varepsilon_i} \right) \\
&= \frac{\lambda_A^{t/2}\lambda_B^{(t+2)/2}\prod_{i=1}^{t-1}(\varepsilon_i)^{t-i}}{\prod_{k=-1}^{(t-2)/2}\left(\lambda_A \sum_{j=k+1}^{t/2}\prod_{i=0}^{2j}\varepsilon_i + \lambda_B \sum_{j=k}^{(t-2)/2}\prod_{i=-1}^{2j+1}\varepsilon_i\right)\prod_{k=0}^{(t-2)/2}\left(\lambda_A \sum_{j=k}^{t/2}\prod_{i=0}^{2j}\varepsilon_i + \lambda_B \sum_{j=k}^{(t-2)/2}\prod_{i=-1}^{2j+1}\varepsilon_i\right)},
\end{aligned}
$$

where $\varepsilon_{-1} = \varepsilon_0 = 1$.

Our proof starts with the observation that the index k on the left-hand side product of the denominator affects the lower bounds of both summations $j = k+1$ and $j = k$, but not the upper ones. This implies that each of the $\frac{t+2}{2}$ terms in the product on the left contains the expression given by the upper bounds $\frac{t}{2}$ and $\frac{t-2}{2}$ respectively, which is $\lambda_A \prod_{i=0}^{t}\varepsilon_i + \lambda_B \prod_{i=-1}^{t-1}\varepsilon_i$. Moreover, all the terms but the last are greater than or equal to this last expression. It follows that

$$
\prod_{k=-1}^{(t-2)/2}\left(\lambda_A \sum_{j=k+1}^{t/2}\prod_{i=0}^{2j}\varepsilon_i + \lambda_B \sum_{j=k}^{(t-2)/2}\prod_{i=-1}^{2j+1}\varepsilon_i\right) \geq \left(\lambda_A \prod_{i=0}^{t}\varepsilon_i + \lambda_B \prod_{i=-1}^{t-1}\varepsilon_i\right)^{(t+2)/2}.
$$

Similarly, we can bound the second term in the denominator by

$$
\prod_{k=0}^{(t-2)/2}\left(\lambda_A \sum_{l=u}^{t/2}\prod_{k=0}^{2l}\varepsilon_k + \lambda_B \sum_{j=k}^{(t-2)/2}\prod_{i=-1}^{2j+1}\varepsilon_i\right) \geq \left(\lambda_A \prod_{i=0}^{t}\varepsilon_i + \lambda_B \prod_{i=-1}^{t-1}\varepsilon_i\right)^{\frac{t}{2}}.
$$

Thus,

$$\prod_{k=-1}^{(t-2)/2}\left(\lambda_A\sum_{j=k+1}^{t/2}\prod_{i=0}^{2j}\varepsilon_i+\lambda_B\sum_{j=k}^{(t-2)/2}\prod_{i=-1}^{2j+1}\varepsilon_i\right)\prod_{k=0}^{(t-2)/2}\left(\lambda_A\sum_{j=k}^{t/2}\prod_{i=0}^{2j}\varepsilon_i+\lambda_B\sum_{j=k}^{(t-2)/2}\prod_{i=-1}^{2j+1}\varepsilon_i\right)$$
$$\geq\left(\lambda_A\prod_{i=0}^{t}\varepsilon_i+\lambda_B\prod_{i=-1}^{t-1}\varepsilon_i\right)^{t+1}$$

It follows that $R_t^A \leq \frac{\lambda_A^{t/2}\lambda_B^{(t+2)/2}\prod_{i=1}^{t-1}(\varepsilon_i)^{t-i}}{\left(\lambda_A\prod_{i=1}^{t}\varepsilon_i+\lambda_B\prod_{i=1}^{t-1}\varepsilon_i\right)^{t+1}}$. Algebra implies

$$R_t^A \leq \frac{\lambda_A^{t/2}\lambda_B^{(t+2)/2}}{(\varepsilon_t\lambda_A+\lambda_B)^{t+1}\prod_{i=1}^{t-1}\varepsilon_i^{i+1}}. \tag{37}$$

Note that the right-hand side term of the inequality trivially goes to zero if any preference value ε_i with i in $\{1,2,\ldots,t\}$ is infinitely big. The intuition is that if voters greatly prefer stability at any given point in time, it is likely that the incumbent will win re-election, if this occurs the alternation cycle is broken. The condition that voters are prone to stability, or are at least indifferent to alternation implies that $\varepsilon_t \geq 1$ for all $t \geq 1$, (37) shows that

$$R_t^A \leq \frac{\lambda_A^{\frac{t}{2}}\lambda_B^{\frac{t+2}{2}}}{(\lambda_A+\lambda_B)^{t+1}}. \tag{38}$$

Let us assume that $\lambda_A = \lambda_B$, so both parties have the same average candidate valence. Substituting λ_B for λ_A into (38) yields $R_t^A \leq \left(\frac{1}{2}\right)^{t+1}$, which converges linearly to zero with rate $\frac{1}{2}$. Explicitly, this means that if candidates are indistinguishable in terms of valence and voters are loyal to the party in power, the probability of perfect alternation is decreasing with time. In other words, we will observe re-election in finite time almost surely.

To study a more general case, we take logarithm in (38) to obtain

$$\ln(R_t^A) \leq \frac{t}{2}\ln(\lambda_A)+\frac{t+2}{2}\ln(\lambda_B)-(t+1)\ln(\lambda_A+\lambda_B). \tag{39}$$

Let us assume that the valence parameter of B is of the form $\lambda_B = c\lambda_A$ with $c > 0$. This form comprises all three possible scenarios: when the candidates of A have a higher average valence than those of B ($c > 1$), when both parties have the same candidate valence ($c = 1$), and when the candidates of B are better suited for office those of A ($c < 1$). We have that

$$(t+1)\ln(\lambda_A) + \frac{t+2}{2}\ln(c) < (t+1)\ln(\lambda_A + \lambda_B) \quad \text{for all } c > 0.$$

This implies

$$\lim_{t\to\infty} \frac{t}{2}\ln(\lambda_A) + \frac{t+2}{2}\ln(\lambda_B) - (t+1)\ln(\lambda_A + \lambda_B) \to -\infty. \tag{40}$$

We conclude from (39) that $R_t^A \leq e^{\frac{t}{2}\ln(\lambda_A)+\frac{t+2}{2}\ln(\lambda_B)-(t+1)\ln(\lambda_A+\lambda_B)}$, hence that

$$\lim_{t\to\infty} R_t^A \leq e^{\lim_{t\to\infty}\frac{t}{2}\ln(\lambda_A)+\frac{t+2}{2}\ln(\lambda_B)-(t+1)\ln(\lambda_A+\lambda_B)},$$

and finally from (40) that $\lim_{t\to\infty} R_t^A \to 0$, which is our assertion. □

Appendix E

```
# R is double the number of periods
R=200
# E's are the uniform advantage vectors
# E1~U(0,1), E2~U(0,4), E3 ~U(0,.2),
E4~U(0,2)
E1=NULL
E2=NULL
E3=NULL
E4=NULL
# P's are vectors of A's partial victories
P1=NULL
P2=NULL
P3=NULL
P4=NULL
# K's average of partial victories vectors
K1=NULL
K2=NULL
K3=NULL
K4=NULL
# X candidate valence vector
X=NULL
# W's indicator functions of victories of
A (0 if A loses, 1 if it wins)
W1=NULL
W2=NULL
W3=NULL
W4=NULL
# N's vector of number of terms in power
N1=NULL
N2=NULL
N3=NULL
N4=NULL
# Sample R exponential variables with
```

```
parameter 1, these are the candidate
valences
for (i in 1:R){
X[i]=rexp(1,rate=1)
}
# Sample the uniform advantages for
R/2 periods
k=R/2
for (i in 1:k)
{
E1[i]=runif(1, min=0, max=1)
E2[i]=runif(1, min=0, max=4)
E3[i]=runif(1, min=0, max=0.2)
E4[i]=runif(1, min=0, max=2)
}
# Obtain the first winning party
if (X[1]<X[2]){
W1[1]=0
W2[1]=0
W3[1]=0
W4[1]=0
} else {
W1[1]=1
W2[1]=1
W3[1]=1
W4[1]=1
}
# Set the first number of terms to 1
N1[1]=1
N2[1]=1
N3[1]=1
N4[1]=1
# Election simulations with
re-election dynamics
for (i in 2:k){
j=2*i
if (N1[i-1]==2){
if (W1[i-1]==1){
if (E1[i]*X[j-1]>X[j]){
W1[i]=1
N1[i]=1
} else if
(E1[i]*X[j-1]<X[j]){
W1[i]=0
N1[i]=1
}
} else if (W1[i-1]==0){
if (X[j-1]<E1[i]*X[j]){
W1[i]=0
N1[i]=1
} else if
(X[j-1]>E1[i]*X[j]){
W1[i]=1
N1[i]=1
}
}
} else if (N1[i-1]==1){
if (W1[i-1]==1){
if (E1[i]*X[j-3]>X[j]){
W1[i]=1
N1[i]=N1[i-1]+1
} else if
(E1[i]*X[j-3]<X[j]){
W1[i]=0
N1[i]=1
}
} else if (W1[i-1]==0){
if (X[j-1]<E1[i]*X[j-2]){
W1[i]=0
N1[i]=N1[i-1]+1
} else if
(X[j-1]>E1[i]*X[j-2]){
W1[i]=1
N1[i]=1
}
}
}
}
for (i in 2:k){
j=2*i
if (N2[i-1]==2){
```

```
if (W2[i-1]==1){
if (E2[i]*X[j-1]>X[j]){
W2[i]=1
N2[i]=1
} else if
(E2[i]*X[j-1]<X[j]){
W2[i]=0
N2[i]=1
}
} else if (W2[i-1]==0){
if (X[j-1]<E2[i]*X[j]){
W2[i]=0
N2[i]=1
} else if
(X[j-1]>E2[i]*X[j]){
W2[i]=1
N2[i]=1
}
}
} else if (N2[i-1]==1){
if (W2[i-1]==1){
if (E2[i]*X[j-3]>X[j]){
W2[i]=1
N2[i]=N2[i-1]+1
} else if
(E2[i]*X[j-3]<X[j]){
W2[i]=0
N2[i]=1
}
} else if (W2[i-1]==0){
if (X[j-1]<E2[i]*X[j-2]){
W2[i]=0
N2[i]=N2[i-1]+1
} else if
(X[j-1]>E2[i]*X[j-2]){
W2[i]=1
N2[i]=1
}
}
}
}
for (i in 2:k){
j=2*i
if (N3[i-1]==2){
if (W3[i-1]==1){
if (E3[i]*X[j-1]>X[j]){
W3[i]=1
N3[i]=1
} else if
(E3[i]*X[j-1]<X[j]){
W3[i]=0
N3[i]=1
}
} else if (W3[i-1]==0){
if (X[j-1]<E3[i]*X[j]){
W3[i]=0
N3[i]=1
} else if
(X[j-1]>E3[i]*X[j]){
W3[i]=1
N3[i]=1
}
}
} else if (N3[i-1]==1){
if (W3[i-1]==1){
if (E3[i]*X[j-3]>X[j]){
W3[i]=1
N3[i]=N3[i-1]+1
} else if
(E3[i]*X[j-3]<X[j]){
W3[i]=0
N3[i]=1
}
} else if (W3[i-1]==0){
if (X[j-1]<E3[i]*X[j-2]){
W3[i]=0
N3[i]=N3[i-1]+1
} else if
(X[j-1]>E3[i]*X[j-2]){
W3[i]=1
N3[i]=1
```

```
}
}
}
}
for (i in 2:k){
j=2*i
if (N4[i-1]==2){
if (W4[i-1]==1){
if (E4[i]*X[j-1]>X[j]){
W4[i]=1
N4[i]=1
} else if
(E4[i]*X[j-1]<X[j]){
W4[i]=0
N4[i]=1
}
} else if (W4[i-1]==0){
if (X[j-1]<E4[i]*X[j]){
W4[i]=0
N4[i]=1
} else if
(X[j-1]>E4[i]*X[j]){
W4[i]=1
N4[i]=1
}
}
} else if (N4[i-1]==1){
if (W4[i-1]==1){
if (E4[i]*X[j-3]>X[j]){
W4[i]=1
N4[i]=N4[i-1]+1
} else if
(E4[i]*X[j-3]<X[j]){
W4[i]=0
N4[i]=1
}
} else if (W4[i-1]==0){
if (X[j-1]<E4[i]*X[j-2]){
W4[i]=0
N4[i]=N4[i-1]+1
} else if
(X[j-1]>E4[i]*X[j-2]){
W4[i]=1
N4[i]=1
}
}
}
}
# Partial number of victories of A
at time t
P1[1]=W1[1]
P2[1]=W2[1]
P3[1]=W3[1]
P4[1]=W4[1]
for (i in 2:k)
{
P1[i]=P1[i-1]+W1[i]
P2[i]=P2[i-1]+W2[i]
P3[i]=P3[i-1]+W3[i]
P4[i]=P4[i-1]+W4[i]
}
# Average number of victories
for (i in 1:k)
{
K1[i]=P1[i]/i
K2[i]=P2[i]/i
K3[i]=P3[i]/i
K4[i]=P4[i]/i
}
# Plotting the average functions
plot(K3, type="l", lwd=2,
lty=1, col="black", xlab="Time",
ylab="Average", xlim=c(0.15, 100),
ylim=c(0.15, 1))
legend("bottomright",legend=c("U(0,0.2)",
"U(0,1)","U(0,2)","U(0,4)"), lty =
c(1,5,4,3), bty ="n")
points(K1, type="l", lwd=2,lty=4,
col="black")
points(K4, type="l", lwd=2, lty=3,
col="black")
```

```
points(K2,   type="l",   lwd=2,   lty=5,   col="black")
```

Empowering of candidates and parties in single term vs re-election schemes

January 22, 2019

The agenda-setting literature aims to elucidate who are the agents that directly or indirectly push policy issues, and the means by which they successfully steer the agenda towards their interests. We highlight two type of theories: one, in which parties see for their own interests, taking advantage of their position in the political control (Green and Mortensen, 2010), and the other in which parties pay attention to the voters' cues, and incorporate their policy concerns into their manifestos (Klüver and Sagarzazu, 2016). Something that stands out is that candidates do not appear to play significant roles in shaping the agenda, at least not directly. In fact, studies of agenda-setting tend to focus more on the role of media in the perception of candidates (Flowers *et al.* 2003, and McCombs, 2004), than on the role of candidates themselves. Kahn (1994) argued that the media agenda seemed to coincide with that of U.S. Senate male candidates, who got higher and more positive coverage in televised advertisements than their female counterparts.

Nevertheless, the figure of candidates should be revisited insofar as it is the elected representatives the ones that implement the policies, giving rise to the classic problem of the principal-agent. As Holcombe and Gwartney (1989) point out, if the party's ability to monitor politicians decreases, then opportunities for deviation in the form of special interest legislation appear. Moreover, there is a strand of literature on the organizational form of parties that claims that parties are shaped and driven by coordinated legislators (Aldrich (1995), and Cox and McCubbins (1993)), which ultimately implies that politicians, and not parties, determine the policy issues. In contrast with this view, the theory of extended party networks (EPN) asserts that parties, along interest groups and activists, selectively support candidates who align with their policy positions, and are likely to advance the party agenda if they are elected (Bawn *et al.*, (2012), Desmarais *et al.* (2015), and Skinner *et al.* (2012)).

Given that the discussion of empowerment is dramatically polarized, we propose a conciliatory model that grants both parties and candidates some degree of legislative leeway, or agenda-setting power, where the share of power of each is negotiated by both sides. In this sense, candidates and parties care about the policy they will implement if they come to power, and thus are *policy-motivated*. We do not explicitly model how the share of power translates into a proper policy in the Downsian ideological space (Downs, 1957), but refer the reader to the work of Serra (2010 and 2018) for a discussion of the impact of candidate valence on policy proposals and polarization.

Our interest is to study the changes in the share by virtue of who is the side that makes the offer. We analyze two situations: one where parties are the proposers, and one where candidates are. The former situation is realistic in strongly ideological parties, where programs are subject to votes by militants, which shapes the primary campaigns, and the latter captures the situation in which candidates raise funds to campaign in primary elections, or where leader's culture matters as much as ideology. Specifically, we propose a dynamic model where parties/candidates offer their preferred candidate/party a share of their ruling power, and the agents that receive the offers pick the option that maximizes their utility. At the end of the process, party-candidate matchings are achieved, and voters decide which pair to support. The first contribution of our work is that we are able to specify the degree to which the parties and the candidates agree on the policies they implement, as opposed to inferring who has the bigger control of the two.

It is worth pointing out that we deal with limits on time in office, explicitly, we conduct our research under the single and two term limit scenarios. The reason is that limits (or their absence) have proven to have behavioral effects on politicians in office, so we wish to investigate if and to what extent, the restriction changes the agenda-setting powers. Our analysis is particularly timely for Mexico, given that in 2014, Mexican President Enrique Peña Nieto announced the end of the ban on re-election for legislators of the Congress, Senate, and mayors, allowing them to serve up to four and two consecutive terms respectively. Carey *et al.* (2006) found that the 1990s adoption of term limits on U.S. legislators changed the priorities of representatives, making their policies less self-interest driven. This result proves Glazer and Wattenberg (1996) conjecture that term limits reduce the incentives of campaigning for re-election through constituency service, and instead direct their efforts towards policy initiatives, it is also one of the arguments of term-limit advocates.

We find that parties get a bigger share of power when they propose to candidates than when they are proposed to. The intuition is that parties take advantage of the preference of candidates and offer as little as possible to secure their preferred match. This also holds for the case where candidates propose first. The implication is that whoever negotiates first, has more power in terms of setting the agenda. More interestingly, moving first has a larger effect on empowering candidates than changing the legal number of terms. Because of this, we expect that the 2014 electoral reform has a minor effect on the political agenda of legislators.

The work is structured as follows. In Section 1 we present the offer-making by parties model under the single term and two term schemes. We obtain and analyze the subgame perfect Nash equilibria. In Section 2 we introduce the models of offer-making by candidates. We compute and interpret the subgame perfect Nash equilibria. Section 3 contains a comparison of the models in terms of convergence to a stationary state, the percentage of elections in which voters choose the candidate with the lowest valence, and the difference in power shares.

10 Offer-making by parties

10.1 Single term model

We consider a dynamic model where two parties, A and B, compete for candidates,[1] offering them a share of their ruling power. Time is discrete, and in each period t in $\{1, 2, 3, \ldots\}$, there is a draft to allocate candidates to parties. The pool consists of two politicians, $c_{1,t}$ and $c_{2,t}$, denoted by c_1 and c_2 whenever there is no confusion. It is assumed that politicians are characterized by their respective valences $\upsilon_{1,t}$ and $\upsilon_{2,t}$, or υ_1 and υ_2, which are random variables that take positive values. The electorate prefers high over low valence candidates.

In considering which match to make, parties and candidates take into account the partisan incumbency advantage ε, a random variable that takes positive values and affects the voter's evaluation of the official candidate . If party A is incumbent and nominates candidate c_1, then the voters regard her valence as $\varepsilon\upsilon_1$ rather than υ_1 alone. A positive advantage $\varepsilon > 1$ enhances the evaluation of the incumbent party's candidate valence, while a negative advantage $0 < \varepsilon < 1$ diminishes it. The evaluation of the challenger's valence remains the same, namely $\varepsilon = 1$ if the party is not in power. We formalize the per period competition between parties for candidates with the following three stage game.

The first move, by nature, determines the candidate valences υ_1 and υ_2, as well as the incumbency advantage factor of party i, ε_i. The realizations are observed by both parties and candidates. In stage 2, parties simultaneously state their most preferred candidate c, and make her an offer of ruling power $0 \leq k \leq 1$. That is, the strategies of party i in $\{A, B\}$ are s_i in $\{c_1, c_2\} \times [0, 1]$, where $s_i = (c, k)$ means that party i makes an offer to candidate c, proposing k percent of the power if they win. Let σ_i be the mixed strategy of i, that is,

[1] Admittedly, parties tend to select their candidates from *disjoint* pools, where potential nominees are aligned with each party's platform. An advantage of maneuvering in a valence dimension independent of ideological issues, is that the same candidate could be appealing for both parties, on the basis of her campaigning skills (See Serra (2011) for a treatment of the effect of valence on the nomination of candidates).

$\sigma_i = (p_{c_r}, 1 - p_{c_r})$, where $0 \leq p_{c_r} \leq 1$ is the probability that party i makes an offer k to candidate c_r, and $0 \leq 1 - p_{c_r} \leq 1$, is the probability that it makes the offer to c_s. In stage 3, candidates pick the offer that maximizes their utility. We denote the strategies of candidate c in $\{c_1, c_2\}$ by q_c, with q_c in $\{O_i^c, O_j^c\}$, where O_i^c means that candidate c accepts the offer of party i. Let σ_c be the mixed strategy of c, so $\sigma_c = (p_i, 1 - p_i)$, where $0 \leq p_i \leq 1$ is the probability that c accepts the offer of party i, and $0 \leq 1 - p_i \leq 1$ is the probability that she accepts the offer of $j \neq i$. If a candidate gets no offer, she is matched with the party rejected by her contender. Once the candidate and party assignments are made, the payoffs are the observed.

The utilities of the winning politician and her party are proportional to the margin of victory.[2] That is, they are positive functions of the distance between the elected candidate's valence and that of the unelected candidate. Let $U_{c_r}(i, k_i)$ denote the utility that candidate c_r gets from running under i's label and having ruling power k_i, and let $U_i(c_r, k_i)$ be the utility that party i gets from running with candidate c_r and granting her power k_i. Then,

$$U_{c_r}(i, k_i) = \begin{cases} 0 & \text{if } \varepsilon_j \upsilon_s > \varepsilon_i \upsilon_r \\ k_i(\varepsilon_i \upsilon_r - \varepsilon_j \upsilon_s) & \text{if } \varepsilon_j \upsilon_s < \varepsilon_i \upsilon_r \end{cases} \tag{41}$$

(where $\varepsilon_j = 1$ if i is the governing party, or $\varepsilon_i = 1$ if j is). The utility of the winning party is the complement of the politician's utility, so $U_i(c_r, k_i) = (1 - k_i)(\varepsilon_i \upsilon_r - \varepsilon_j \upsilon_s)$. The extensive-form game is portrayed in Figure 8.

Theorem 4. *Let i in $\{A, B\}$ be the governing party of period t, and let $0 < \varepsilon_i \leq 1$ be its incumbency advantage. The requirement on candidates c_r and c_s in $\{c_1, c_2\}$ is that $\upsilon_r < \upsilon_s$.*

1. *If $\varepsilon_i \leq \frac{\upsilon_r}{\upsilon_s}$, then subgame perfect Nash equilibria are $s_i^* = \sigma_i$, where σ_i is any mixed strategy with k in $[0, 1]$, $s_j^* = (c_s, 0)$ with $j \neq i$, $q_{c_r}^* = \sigma_{c_r}$, and $q_{c_s}^* = O_j^{c_s}$.*

[2] This functional form corresponds to a proportional representation system, where the number of seats that each party gets, and indirectly its control over policy issues, is proportional to the votes the party gets.

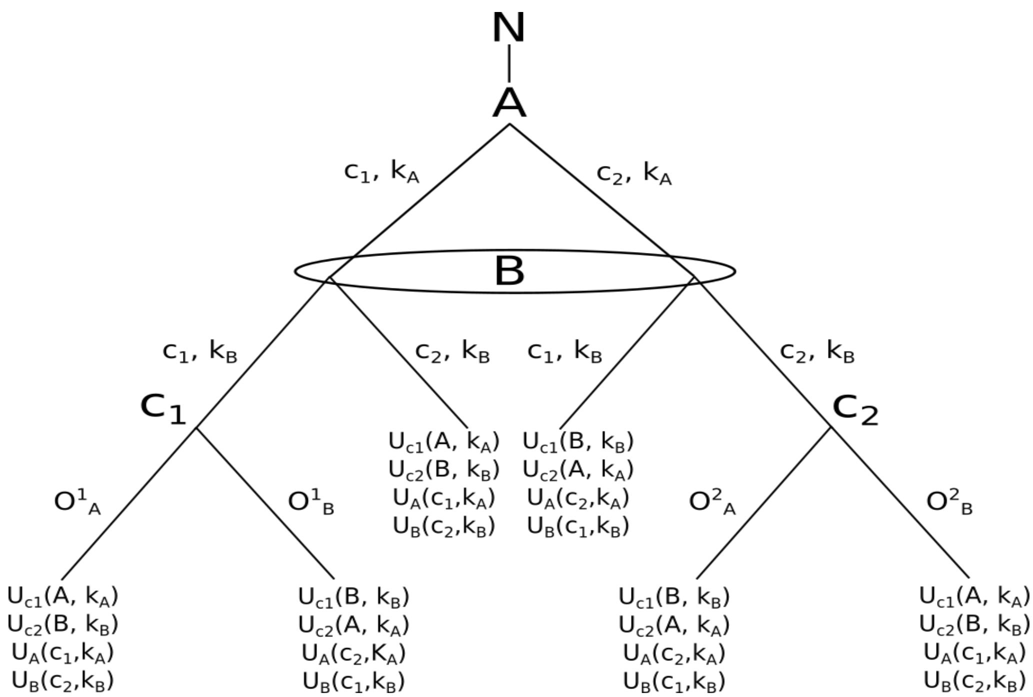

Figure 8: Single term game with offers by parties

2. *If* $\frac{\upsilon_r}{\upsilon_s} < \varepsilon_i < 1$, *then subgame perfect Nash equilibrium is unique and is* $s_i^* = (c_s, 1)$, $s_j^* = \left(c_s, \frac{\varepsilon_i \upsilon_s - \upsilon_r}{\upsilon_s - \varepsilon_i \upsilon_r}\right)$, $q_{c_r}^* = \sigma_{c_r}$, *and* $q_{c_s}^* = O_j^{c_s}$.

3. *If* $\varepsilon_i = 1$, *then subgame perfect Nash equilibria are* $s_i^* = (c_s, 1)$, $s_j^* = (c_s, 1)$, $q_{c_r}^* = \sigma_{c_r}$ *and* $q_{c_s}^* = \sigma_{c_s}$, *where* σ_{c_r} *and* σ_{c_s} *are any mixed strategies.*

The theorem shows that when the partisan incumbency advantage is negative, and such that no candidate that runs under the governing party's label stands a chance of winning, that is when $\varepsilon_i < \frac{\upsilon_r}{\upsilon_s}$, then any offer from the opposition k_j is preferred over any offer from the governing party k_i. The opposition takes advantage of this, and in equilibrium offers no power to the best candidate c_s, knowing its offer will be accepted with certainty. The incumbent party is aware that none of its offers is appealing to either candidate, so it randomly picks a k_i in $[0, 1]$ and offers this share to the best candidate as well. Evidently, the number of possible offers from i is infinite, and therefore the equilibrium is not unique.

If however, the advantage is negative but its effect is not enough to prevent candidates from winning, that is if $\frac{\upsilon_r}{\upsilon_s} \leq \varepsilon_i < 1$, then parties compete for the candidacy of the highest valued politician through their offers. The assertion is that the incumbent party grants her all the ruling power in order to secure the match, while the opposition offers the bare minimum amount k_j that makes c_s indifferent between accepting $k_i = 1$ or k_j. It is a simple matter to check that there is no other pair of best responses on the parties' side, so the equilibrium is unique. We refer the reader to Appendix F for equilibria when the partisan advantage is positive; it suffices to say that the reasoning is analogous to the described above.

Finally, when the partisan incumbency advantage is negligible, so $\varepsilon_i = 1$, the best politician is guaranteed to win under the label of either party, this implies that parties compete in offers for her candidacy. The parties' best response is to offer c_s the maximum amount of ruling power.

Proof. Throughout the proof, we assume that the valences of c_1 and c_2 are such that $\upsilon_1 < \upsilon_2$, and that party A is incumbent at time t. This involves no loss of generality.
Given the sequential nature of the game, we use backward induction to compute the subgame perfect Nash equilibria. Since the third stage corresponds to the choice-making of candidates, we restrict our attention to the case where a single candidate receives offers from both parties. From the definition of utilities given in (41) we have

$$U_{c_1}(A, k_A) = \begin{cases} 0 & \text{if } \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1} \\ k_A(\varepsilon_A \upsilon_1 - \upsilon_2) & \text{if } \varepsilon_A > \frac{\upsilon_2}{\upsilon_1} \end{cases} \quad U_{c_1}(B, k_B) = \begin{cases} k_B(\upsilon_1 - \varepsilon_A \upsilon_2) & \text{if } \varepsilon_A < \frac{\upsilon_1}{\upsilon_2} \\ 0 & \text{if } \varepsilon_A \geq \frac{\upsilon_1}{\upsilon_2} \end{cases}, \tag{42}$$

$$U_{c_2}(A, k_A) = \begin{cases} 0 & \text{if } \varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2} \\ k_A(\varepsilon_A \upsilon_2 - \upsilon_1) & \text{if } \varepsilon_A > \frac{\upsilon_1}{\upsilon_2} \end{cases} \quad U_{c_2}(B, k_B) = \begin{cases} k_B(\upsilon_2 - \varepsilon_A \upsilon_1) & \text{if } \varepsilon_A < \frac{\upsilon_2}{\upsilon_1} \\ 0 & \text{if } \varepsilon_A \geq \frac{\upsilon_2}{\upsilon_1} \end{cases}, \tag{43}$$

Preference	$\succeq_{c_2}$		
$\frac{k_A}{k_B}$	$< c$	$= c$	$> c$
Most	B	A/B	A
Least	A	B/A	B

Table 2: Preference order with negative incumbency advantage

this leads to the following observation.

Step 1. *If $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$, then $U_{c_1}(A, k_A) < U_{c_1}(B, k_B)$ for all $k_A \geq 0$ and $k_B > 0$. If $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$, then $U_{c_2}(A, k_A) < U_{c_2}(B, k_B)$ for all $k_A \geq 0$ and $k_B > 0$. If $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1$, then $U_{c_1}(A, k_A) = U_{c_1}(B, k_B) = 0$ for all $k_A, k_B \geq 0$. If $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$, then $U_{c_2}(A, k_A) = k_A(\varepsilon_A \upsilon_2 - \upsilon_1)$ and $U_{c_2}(B, k_B) = k_B(\upsilon_2 - \varepsilon_A \upsilon_1)$.*

This allows one to see that when $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$, both candidates have a dominant strategy, which is to accept any positive offer from B. It also implies that if $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1$, then c_1 is indifferent between running under A or B. If $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$, the preference of c_2 is determined by the offers she gets. The following table shows the preference of c_2 when $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_i \leq 1$, as a function of the ratio of offers $\frac{k_A}{k_B}$. We define $c = \frac{\upsilon_2 - \varepsilon_A \upsilon_1}{\varepsilon_A \upsilon_2 - \upsilon_1}$ to be the indifference curve of offers, it is obtained by equalizing $U_{c_2}(A, k_A)$ and $U_{c_2}(B, k_B)$, and clearing the ratio $\frac{k_A}{k_B}$. The information is displayed graphically in Figure 9.

The preceding observation and the interpretation of Figure 9, lead to the following step.

Step 2. *We define $c := \frac{\upsilon_2 - \varepsilon_A \upsilon_1}{\varepsilon_A \upsilon_2 - \upsilon_1}$. If $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$, then c_1 has the dominant strategy to accept any positive offer from B. If $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$, then c_2 has the dominant strategy to accept any positive offer from B. If $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1$, then c_1 is indifferent between accepting any offer from A or B. If $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$ and $\frac{k_A}{k_B} < c$, then c_2 has the dominant strategy to accept B's offer. If $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$ and $\frac{k_A}{k_B} > c$, then c_2 has the dominant strategy to accept A's offer. Lastly, if $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$ and $\frac{k_A}{k_B} = c$, then c_2 is indifferent between accepting the offer of A and B. Thus, the strategies of c_1 and c_2, q_{c_1} and q_{c_2}, are*

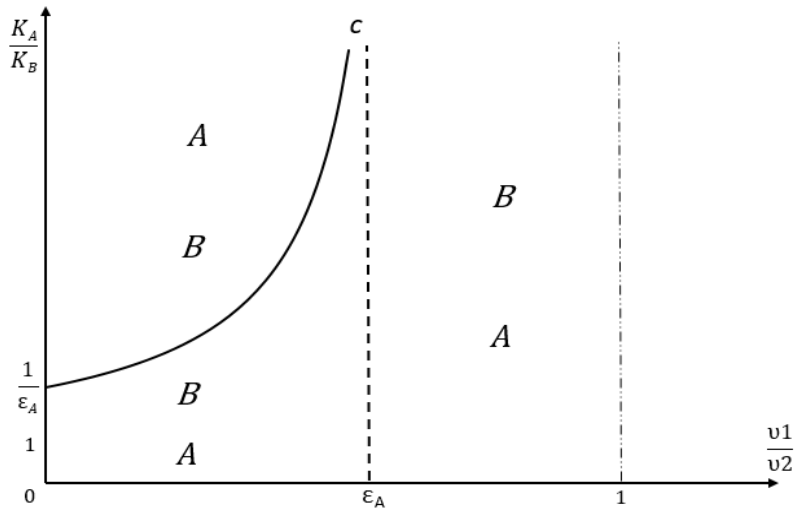

Figure 9: Preferences of c_2 with negative incumbency advantage

- $q_{c_1}(s_A, s_B) = O_B^{c_1}$ for all $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$ and $\frac{k_A}{k_B} \geq 0$
- $q_{c_1}(s_A, s_B) = \sigma_{c_1}$ for all $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1$ and $\frac{k_A}{k_B} \geq 0$
- $q_{c_2}(s_A, s_B) = O_B^{c_2}$ for all $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$ and $\frac{k_A}{k_B} \geq 0$
- $q_{c_2}(s_A, s_B) = O_B^{c_2}$ for all $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$ and $\frac{k_A}{k_B} < c$
- $q_{c_2}(s_A, s_B) = O_A^{c_2}$ for all $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$ and $\frac{k_A}{k_B} > c$
- $q_{c_2}(s_A, s_B) = \sigma_{c_2}$ for all $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$ and $\frac{k_A}{k_B} = c$.

Given the best responses of candidates, our next concern is the decision-making of parties in stage 2. Here again, we turn to the definition of utilities given in (41), plugging the dominant strategies of candidates established at step 2, we find that the only strictly positive utilities are

$$
\begin{aligned}
U_A((c_2,k_A) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) &= (1-k_A)(\varepsilon_A \upsilon_2 - \upsilon_1) \quad \text{if } \frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1 \text{ and } \frac{k_A}{k_B} \geq c \\
U_B((c_1,k_B) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) &= (1-k_B)(\upsilon_1 - \varepsilon_A \upsilon_2) \quad \text{if } \varepsilon_A < \frac{\upsilon_1}{\upsilon_2} \text{ and } \frac{k_A}{k_B} \geq 0 \\
U_B((c_2,k_B) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) &= (1-k_B)(\upsilon_2 - \varepsilon_A \upsilon_1) \quad \text{if } \varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2} \text{ and } \frac{k_A}{k_B} \geq 0 \\
U_B((c_2,k_B) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) &= (1-k_B)(\upsilon_2 - \varepsilon_A \upsilon_1) \quad \text{if } \frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1 \text{ and } \frac{k_A}{k_B} \leq c.
\end{aligned}
$$

Note that $U_A((c_1,k_A) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) < U_A((c_2,k_A) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B))$ only if $\varepsilon_A > \frac{\upsilon_1}{\upsilon_2}$ and $\frac{k_A}{k_B} > c$, and $U_B((c_1,k_B) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) < U_B((c_2,k_B) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B))$ for all $\frac{k_A}{k_B} \geq 0$ if $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$, or for all $\frac{k_A}{k_B} > c$ if $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$.
Therefore, if the negative advantage is such that no candidate associated with A has a chance of winning the election, that is if $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$, then A is indifferent between making any offer k in $[0,1]$ to c_1 or c_2, as all of them will be rejected. It is only when $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$ that A has incentives to compete for c_2's candidacy against B's offer, for if $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$ and $\frac{k_A}{k_B} > c$, then it gets a positive utility. Given that A is guaranteed to lose the election if $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$, B is the preferred party of both candidates, and they accept any positive offer k_B from it. Since $U_B(c_1,k_B) < U_B(c_2,k_B)$ when $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$, B makes its offer to c_2. If $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$, then B also prefers to run with c_2, nevertheless, it knows it competes against A's offer k_A. Thus, for B to achieve a positive utility, it must be the case that it offers k_B such that $\frac{k_A}{k_B} < c$. This argument leads to the following strategies of parties.

Step 3. *Let* $c = \frac{\upsilon_2 - \varepsilon_A \upsilon_1}{\varepsilon_A \upsilon_2 - \upsilon_1}$. *If* $0 < \varepsilon_A \leq 1$, *the strategies of parties* A *and* B, s_A *and* s_B, *are*

- $s_A(k_A,k_B) = \sigma_A$ *if* $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$
- $s_A(k_A,k_B) = (c_2,k_A)$ *with* k_A *such that* $k_A \geq ck_B$ *if* $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$
- $s_B(k_A,k_B) = (c_2,k_B)$ *with* k_B *in* $(0,1]$ *if* $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$
- $s_B(k_A,k_B) = (c_2,k_B)$ *with* k_B *such that* $k_A \leq ck_B$ *if* $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$.

Our last step concerns the utility maximization of parties. To this end, we analyze their best responses.

Step 4. *First note that if $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$, then $s_A(k_A, k_B^*)$ is not a function of k_B^*, nor is $s_B(k_A^*, k_B)$ of k_A^*. This implies that each party maximizes its utility with respect to its own offer. Thus,*

1. *$s_A^* = \sigma_A$ if $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$*

2. *$s_B^* = (c_2, 0)$ if $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$.*

If $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$, then $s_A(k_A, k_B^*)$ and $s_B(k_A^*, k_B)$ are indeed functions of the strategy of the other party. Holding them as given, and reacting accordingly leads to the following.

1. $s_A^* = (c_2, 1)$ if $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$

2. $s_B^* = \left(c_2, \frac{\varepsilon_A \upsilon_2 - \upsilon_1}{\upsilon_2 - \varepsilon_A \upsilon_1}\right)$ if $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$.

Incidentally, we have proved that when $\varepsilon_A = 1$, the best responses of A and B are to offer $k_A = k_B = 1$ to c_2. We provide the intuition behind this result. Note that if party i in $\{A, B\}$ were to offer $k_i = k$ with $0 \leq k < 1$, then the opposition would secure c_2 by offering $k_j = k + \epsilon$, where ϵ is an infinitesimally small quantity. It is clear that i would increase its offer to $k_i = k_j + \epsilon$. Repeating this argument leads to the desired result.

Steps 2 and 4 lead to Theorem 4. $\square$

10.2 Two term model

We consider a dynamic model with two parties, A and B. We assume that time is discrete, and that each period t in $\{1, 2, 3, \ldots\}$, candidates are matched to parties. The difference with respect to the previous model is that politicians are now allowed to spend two terms in office. This implies that incumbents can run for re-election if they have been in office one term, it also means that incumbents vacate their position after their second term, and that in this period an open election is held. We emphasize that by definition, an open election has no personal incumbency advantages, but it does involve a partisan advantage. In this regard, in the period following a re-election, there is a game like the one described in the single term model. Since this case has been discussed in section 10.1, we restrict our attention to the case where an incumbent runs for re-election.

We assume that the incumbent runs for the governing party, and that the opposition has to pick its challenger. The opposition is myopic in the sense that, in electing its candidate, it only considers the utility of the current period. That is, it disregards the fact that if the challenger were to win the election, she would run for a second term. The pool consists of two politicians, $c_{1,t}$ and $c_{2,t}$, denoted by c_1 and c_2 whenever there is no confusion. It is assumed that politicians are characterized by their respective valences $\upsilon_{1,t}$ and $\upsilon_{2,t}$, or υ_1 and υ_2, which are random variables that take positive values. The electorate prefers high over low valence candidates. We formalize the opposition's choice for the challenger with the following three stage game.

In the first stage, nature determines the candidate valences υ_1 and υ_2. The realizations are observed by both parties and candidates. In stage 2, the opposition states its most preferred candidate c, and makes her an offer of ruling power k in $[0, 1]$. That is, the strategies of the opposition j in $\{A, B\}$ are s_j in $\{c_1, c_2\} \times [0, 1]$, where $s_j = (c, k)$ means that party j makes an offer to candidate c, proposing k percent of the power if they win. Let σ_j be the mixed strategy of party j, that is, $\sigma_j = (p_{c_r}, 1 - p_{c_r})$ where $0 \leq p_{c_r} \leq 1$ is the probability that j makes an offer k to candidate c_r, and $0 \leq 1 - p_{c_r} \leq 1$ is the probability that it makes the offer to c_s. In stage 3, the preferred candidate c in $\{c_1, c_2\}$ evaluates the offer and decides whether to accept it or to reject it. So, her strategies, denoted by q_c, are q_c in $\{A, R\}$, where A stands for accepting, and R stands for rejecting the offer. Let σ_c be the mixed strategy of c, that is, $\sigma_c = (p_A, p_R)$, where $0 \leq p_A \leq 1$ is the probability that candidate c accepts the offer, and $0 \leq p_R = 1 - p_A \leq 1$ is the probability that she rejects it. We make the assumption that if indifferent between taking the offer and not running, she decides to run for office under the opposition's label. Once the candidate and party assignments are made, the payoffs are observed.

We continue to assume that the utilities of the winning politician and her party are proportional to the margin of victory. Thus, they are by definition, the same as those given in (41). The extensive-form game is portrayed in Figure 10.

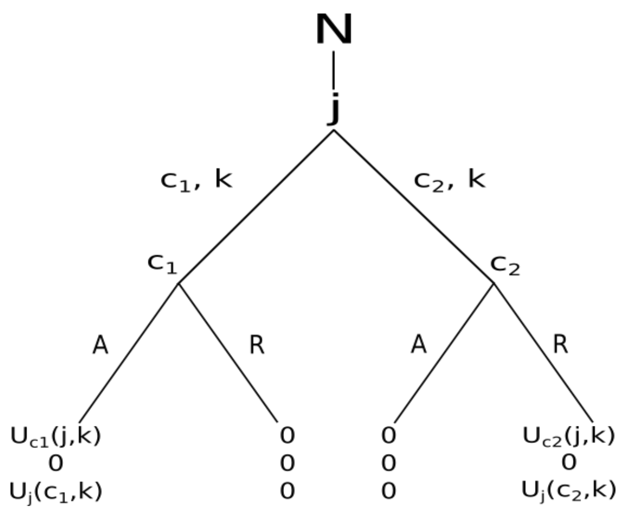

Figure 10: Two term game with offers by parties

Theorem 5. *Let i in $\{A, B\}$ be the governing party of period t, and let $\varepsilon_i > 0$ be its incumbency advantage. Thus, $j \neq i$ is the opposition. The candidates c_r and c_s in $\{c_1, c_2\}$ are such that $\upsilon_r < \upsilon_s$. The subgame perfect Nash equilibria are $s_j^* = (c_s, 0)$, $q_{c_r}^* = \sigma_{c_r}$, where σ_{c_r} is any mixed strategy, and $q_{c_s}^* = A$.*

Proof. Throughout the proof, we assume that the valences of c_1 and c_2 are such that $\upsilon_1 < \upsilon_2$, and that party A is incumbent at time t. We also assume that the valence evaluation of the incumbent is $\varepsilon_A \upsilon_A$, with $\varepsilon_A > 0$. This involves no loss of generality.

Given the sequential nature of the game, we use backward induction to compute the subgame perfect Nash equilibrium. Since the third stage corresponds to the choice-making of candidates, we restrict our attention to the case where a politician receives an offer from the opposition.

From the definition of utilities given in (41) we have

$$U_{c_1}(A,k) = \begin{cases} 0 & \text{if } \frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A \\ k(\upsilon_1 - \varepsilon_A \upsilon_A) & \text{if } \frac{\upsilon_1}{\upsilon_A} > \varepsilon_A \end{cases} \qquad U_{c_1}(R,k) = 0,$$

$$U_{c_2}(A,k) = \begin{cases} 0 & \text{if } \frac{\upsilon_2}{\upsilon_A} \leq \varepsilon_A \\ k(\upsilon_2 - \varepsilon_A \upsilon_A) & \text{if } \frac{\upsilon_2}{\upsilon_A} > \varepsilon_A \end{cases} \qquad U_{c_2}(R,k) = 0.$$

Step 1. *From the above utilities we have that $U_c(R) \leq U_c(A)$ for all c in $\{c_1, c_2\}$, $\varepsilon_A > 0$, and $k \geq 0$. In other words, accepting is a weakly dominant strategy for both parties.*

This, in conjunction with the fact that indifferent candidates accept the candidacy of the opposition, lead to the following strategies.

Step 2. *If candidate c receives an offer $k > 0$ from party B, then*

- *$q_c(k) = A$ for c in $\{c_1, c_2\}$.*

Given these strategies, our next concern is the decision-making of the opposition. Here again, we turn to the definition of utilities given in (41), plugging the dominant strategies of candidates established at step 2, we find that the only strictly positive utilities are

$$\begin{aligned} U_B((c_1,k) \mid q_{c_1}(k), q_{c_2}(k)) &= (1-k)(\upsilon_1 - \varepsilon_A \upsilon_A) \quad \text{if } \frac{\upsilon_1}{\upsilon_A} > \varepsilon_A \text{ and } k \geq 0 \\ U_B((c_2,k) \mid q_{c_1}(k), q_{c_2}(k)) &= (1-k)(\upsilon_2 - \varepsilon_A \upsilon_A) \quad \text{if } \frac{\upsilon_2}{\upsilon_A} > \varepsilon_A \text{ and } k \geq 0. \end{aligned}$$

By the assumption that $\upsilon_1 < \upsilon_2$, we have $U_B((c_1,k) \mid q_{c_1}(k), q_{c_2}(k)) \leq U_B((c_2,k) \mid q_{c_1}(k), q_{c_2}(k))$ for all $k \geq 0$. This leads to the following strategy.

Step 3. *Given that $U_B(c_1,k) \leq U_B(c_2,k)$ for all $k \geq 0$, making an offer to the highest valued candidate c_2 is a weakly dominant strategy for B. That is,*

- *$s_B(c,k) = (c_2,k)$ for all $k \geq 0$ and $\varepsilon_A > 0$.*

The last step concerns the utility maximization of B.

Step 4. *By definition of utility, B maximizes its payoff by offering the least amount of power to its candidate. Thus,*

1. $s_B^*(c,k) = (c_2, 0)$ *for all* $\varepsilon_A > 0$.

Steps 2 and 4 lead to Theorem (5). □

11 Offer-making by candidates

11.1 Single term model

We consider a dynamic model where two politicians c_1 and c_2, present their candidacy to a political party, A or B, offering them a share of ruling power. Time is discrete, and each period t in $\{1, 2, 3, \ldots\}$, politicians are matched to parties. The two politicians are characterized by their respective valences, υ_1 and υ_2, which are random variables that take positive values. The electorate prefers high over low valence candidates.

In considering which match to make, politicians and parties take into account the partisan incumbency advantage ε, a random variable that takes positive values and affects the voter's evaluation of the official candidate just as described in section 10.1. We formalize the per period competition between politicians for a given candidacy with the following three stage game.

In stage 1, nature determines the valences of politicians c_1 and c_2, υ_1 and υ_2, as well as the incumbency advantage of party i, ε_i. The realizations are observed by both parties and politicians. In stage 2, politicians simultaneously offer their most preferred party i in $\{A, B\}$ a share $1-k$ in $[0,1]$ of the power if they win. So, the strategies of politician c in $\{c_1, c_2\}$, denoted by q_c, are $q_c = (i,k)$ in $\{A, B\} \times [0,1]$, where $q_c = (i,k)$ means that politician c offers party i, $1-k$ percent of power, or alternatively, offers to keep k percent to herself. Let σ_c be the mixed strategy of c, that is, $\sigma_c = (p_i, 1-p_i)$, where $0 \leq p_i \leq 1$ is the probability that politician c makes an offer k to party i, and $0 \leq 1 - p_i \leq 1$ is the probability that c makes the offer to party $j \neq i$. In stage 3, parties pick the offer that maximizes their utility. We denote the strategies of party i in $\{A, B\}$ by s_i, with s_i in $\{O^i_{c_r}, O^i_{c_s}\}$, where $O^i_{c_r}$ means that party i accepts the offer of candidate c_r. Let σ_i be the mixed strategy of party i, so $\sigma_i = (p_{c_r}, 1 - p_{c_r})$, where $0 \leq p_{c_r} \leq 1$ is the probability that party i accepts the offer k from politician c_r. If a party gets no candidacy, it is matched with the politician rejected by the other party. Once the politician and party assignments are made, the payoffs are observed.

The utilities of the winning politician and her party are proportional to the margin of victory. Let $U_{c_r}(i, k_r)$ denote the utility that candidate c_r gets from running under i's label and having ruling power k_r, and let $U_i(c_r, k_r)$ be the utility that party i gets from running with politician c_r, and keeping $1 - k_r$ of power. Then,

$$U_{c_r}(i, k_i) = \begin{cases} 0 & \text{if } \varepsilon_j \upsilon_s > \varepsilon_i \upsilon_r \\ k_i(\varepsilon_i \upsilon_r - \varepsilon_j \upsilon_s) & \text{if } \varepsilon_j \upsilon_s < \varepsilon_i \upsilon_r \end{cases} \tag{44}$$

(where $\varepsilon_j = 1$ if i is the governing party, or $\varepsilon_i = 1$ if j is). The utility of the winning party is the complement of the politician's utility, so $U_i(c_r, k_i) = (1 - k_i)(\varepsilon_i \upsilon_r - \varepsilon_j \upsilon_s)$. The extensive-form game is portrayed in Figure 11.

The following result may be proved in much the same way as Theorem 1.

Theorem 6. *Let i in $\{A, B\}$ be the governing party of period t, and let $\varepsilon_i \leq 1$ be its incumbency advantage. The requirement on candidates c_r and c_s in $\{c_1, c_2\}$ is that $\upsilon_r < \upsilon_s$.*

1. *If $\frac{\upsilon_r}{\upsilon_s} \leq \varepsilon_i$, then subgame perfect Nash equilibria are $q^*_{c_r} = \sigma_{c_r}$, $q^*_{c_s} = (B, 1)$, $s^*_i = \sigma_i$, and $s^*_j = O^j_{c_s}$.*
2. *If $\varepsilon_A < \frac{\upsilon_r}{\upsilon_s} \leq 1$, then subgame perfect Nash equilibrium is $q^*_{c_r} = (B, 0)$, $q^*_{c_s} = \left(B, 1 - \frac{\upsilon_r - \varepsilon_i \upsilon_s}{\upsilon_s - \varepsilon_i \upsilon_r}\right)$, $s^*_i = \sigma_i$, where σ_i is any mixed strategy, and $s^*_j = O^j_{c_s}$.*

We refer the reader to Appendix G for equilibria when the partisan advantage is negative. The reasoning is analogous to the described above.

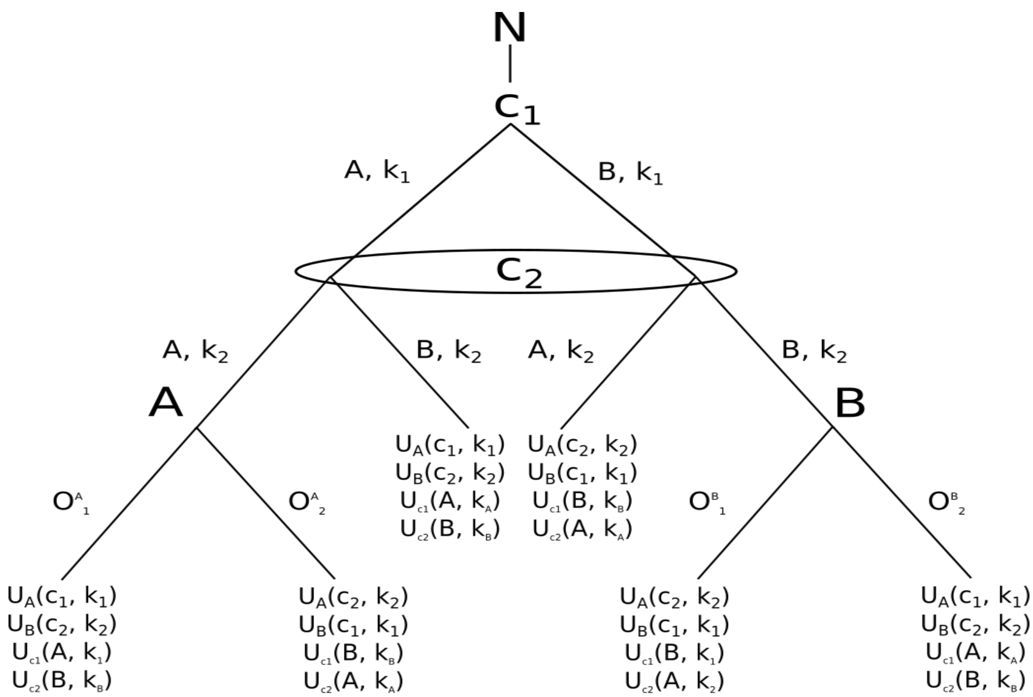

Figure 11: Single term game with offers by candidates

Proof. Throughout the proof, we assume that the valences of c_1 and c_2 are such that $\upsilon_1 < \upsilon_2$, and that party A is incumbent at time t, this involves no loss of generality.

Given the sequential nature of the game, we use backward induction to compute the subgame perfect Nash equilibria. Since the third state corresponds to the choice-making of parties, we restrict our attention to the case where a single party receives offers from both politicians. From the definition of utilities given in (41) we have

$$U_A(c_1,k_1) = \begin{cases} 0 & \text{if } \varepsilon_A \le \frac{\upsilon_2}{\upsilon_1} \\ (1-k_1)(\varepsilon_A\upsilon_1 - \upsilon_2) & \text{if } \varepsilon_A > \frac{\upsilon_2}{\upsilon_1} \end{cases} \quad U_A(c_2,k_2) = \begin{cases} 0 & \text{if } \varepsilon_A \le \frac{\upsilon_1}{\upsilon_2} \\ (1-k_2)(\varepsilon_A\upsilon_2 - \upsilon_1) & \text{if } \varepsilon_A > \frac{\upsilon_1}{\upsilon_2} \end{cases}, \tag{45}$$

Preference	$\succeq_B$		
$\frac{m_1}{m_2}$	$< c$	$= c$	$> c$
Most	c_2	c_1/c_2	c_1
Least	c_1	c_2/c_1	c_2

Table 3: Preference order with negative incumbency advantage

$$U_B(c_1,k_1) = \begin{cases} (1-k_1)(\upsilon_1 - \varepsilon_A \upsilon_2) & \text{if } \varepsilon_A < \frac{\upsilon_1}{\upsilon_2} \\ 0 & \text{if } \varepsilon_A \geq \frac{\upsilon_1}{\upsilon_2} \end{cases} \quad U_B(c_2,k_2) = \begin{cases} (1-k_2)(\upsilon_2 - \varepsilon_A \upsilon_1) & \text{if } \varepsilon_A < \frac{\upsilon_2}{\upsilon_1} \\ 0 & \text{if } \varepsilon_A \geq \frac{\upsilon_2}{\upsilon_1} \end{cases}, \tag{46}$$

Step 1. *From the above utilities, we have that if $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$, then $U_A(c_1,k_1) = U_A(c_2,k_2)$ for all $k_1, k_2 \geq 0$. If $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$, then $U_A(c_1,k_1) < U_A(c_2,k_2)$ for all $k_1 \geq 0$ and $k_2 > 0$. If $\varepsilon_A \geq \frac{\upsilon_1}{\upsilon_2}$, then $U_B(c_1,k_1) < U_B(c_2,k_2)$ for all $k_1 \geq 0$ and $k_2 > 0$. If $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$, then $U_B(c_1,k_1) = (1-k_1)(\upsilon_1 - \varepsilon_A\upsilon_2)$, and $U_B(c_2,k_2) = (1-k_2)(\upsilon_2 - \varepsilon_A\upsilon_1)$.*

In other words, if $\varepsilon_A \leq 1$, then accepting any offer $k_2 > 0$ from c_2 is a weakly dominant strategy for A. If $\varepsilon_A \geq \frac{\upsilon_1}{\upsilon_2}$, then accepting any offer $k_2 > 0$ from c_2 is a dominant strategy of B. If $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$, then the preference of B is determined by the offers she gets. Table 3 shows the preference of B when $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$, as a function of the ratio $\frac{m_1}{m_2}$, where $m_1 = 1 - k_1$ is the percentage of power B gets if c_1 wins the election, and $m_2 = 1 - k_2$ is the percentage it gets if c_2 does. We define $c = \frac{\upsilon_2 - \varepsilon_A \upsilon_1}{\upsilon_1 - \varepsilon_A \upsilon_2}$ to be the indifference curve of power offers for B, it is obtained by equalizing $U_B(c_1,k_1)$ and $U_B(c_2,k_2)$, and clearing the ratio $\frac{m_1}{m_2}$. The information is displayed graphically in Figure 12.

The preceding observation and the interpretation of Figure 12, lead to the following step.

Step 2. *We define $c := \frac{\upsilon_2 - \varepsilon_A \upsilon_1}{\upsilon_1 - \varepsilon_A \upsilon_2}$.*

- $s_A(q_{c_1}, q_{c_2}) = \sigma_{c_2}$ *for all $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$, $k_1 \geq 0$, and $k_2 \geq 0$*

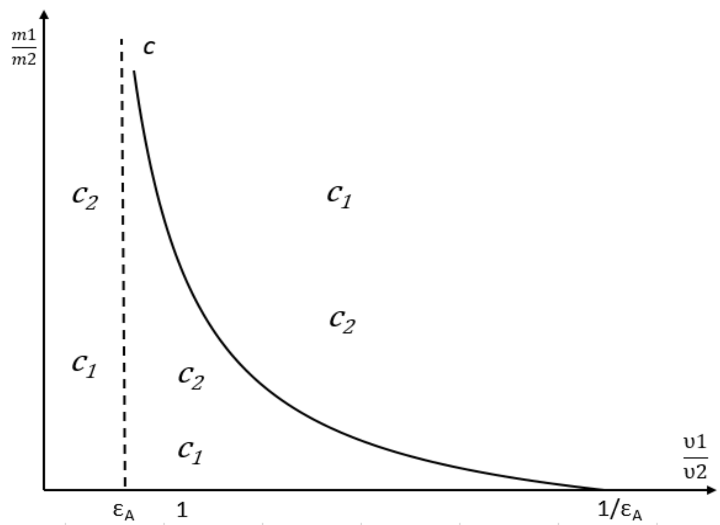

Figure 12: Preferences of B with negative incumbency advantage

- $s_A(q_{c_1}, q_{c_2}) = O^A_{c_2}$ *for all* $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1,\ k_1 \geq 0,$ *and* $k_2 > 0$
- $s_B(q_{c_1}, q_{c_2}) = O^B_{c_2}$ *for all* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2} \leq 1,$ *and* $\frac{1-k_1}{1-k_2} < c$
- $s_B(q_{c_1}, q_{c_2}) = O^B_{c_1}$ *for all* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2} \leq 1,$ *and* $\frac{1-k_1}{1-k_2} > c$
- $s_B(q_{c_1}, q_{c_2}) = \sigma_B$ *for all* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2} \leq 1,$ *and* $\frac{1-k_1}{1-k_2} = c$
- $s_B(q_{c_1}, q_{c_2}) = O^B_{c_2}$ *for all* $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1,\ k_1 \geq 0,$ *and* $k_2 > 0$

Given the best responses of parties, our next concern is the decision-making of candidates in stage 2. We turn to the definition of utilities given in (44), plugging the dominant strategies of parties established at step 2, we find that the only strictly positive utilities are

$$\begin{aligned}
U_{c_1}((B,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2)) &= k_1(\upsilon_1 - \varepsilon_A\upsilon_2) \quad \text{if } \varepsilon_A < \frac{\upsilon_1}{\upsilon_2} \text{ and } \frac{1-k_1}{1-k_2} \geq c \\
U_{c_2}((A,k_2) \mid s_A(k_1,k_2), s_B(k_1,k_2)) &= k_2(\varepsilon_A\upsilon_2 - \upsilon_1) \quad \text{if } \frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1 \text{ and } \frac{1-k_1}{1-k_2} \geq 0 \\
U_{c_2}((B,k_2) \mid s_A(k_1,k_2), s_B(k_1,k_2)) &= k_2(\upsilon_2 - \varepsilon_A\upsilon_1) \quad \text{if } \varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2} \text{ and } \frac{1-k_1}{1-k_2} \leq c \\
U_{c_2}((B,k_2) \mid s_A(k_1,k_2), s_B(k_1,k_2)) &= k_2(\upsilon_2 - \varepsilon_A\upsilon_1) \quad \text{if } \frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1 \text{ and } \frac{1-k_1}{1-k_2} \geq 0.
\end{aligned}$$

Note that $U_{c_1}((A,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2)) < U_{c_1}((B,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2))$ if $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$ and $\frac{1-k_1}{1-k_2} \geq c$, and $U_{c_2}((A,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2)) < U_{c_2}((B,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2))$ if $\varepsilon_A \leq \frac{\upsilon_1}{\upsilon_2}$ and $\frac{1-k_1}{1-k_2} \leq c$, or if $\frac{\upsilon_1}{\upsilon_2} < \varepsilon_A \leq 1$ and $\frac{1-k_1}{1-k_2} \geq 0$. This leads to the following strategies of candidates.

Step 3. *Let $c = \frac{\upsilon_2 - \varepsilon_A \upsilon_1}{\upsilon_1 - \varepsilon_A \upsilon_2}$, the strategies of candidates c_1 and c_2, q_{c_1} and q_{c_2}, are*

- $q_{c_1}(k_1,k_2) = (B,k_1)$ with k_1 such that $1-k_1 \geq c(1-k_2)$ if $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$
- $q_{c_1}(k_1,k_2) = \sigma_{c_1}$ with $k_1 \geq 0$ if $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1$
- $q_{c_2}(k_1,k_2) = (B,k_2)$ with k_2 such that and $1-k_1 \leq c(1-k_2)$ if $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$
- $q_{c_2}(k_1,k_2) = (B,k_2)$ with k_2 in $(0,1]$ if $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1$

Our last step concerns the utility maximization of candidates. To this end, we analyze their best responses.

Step 4. *First note that if $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1$, then $q_{c_1}(k_1,k_2^*)$ is not a function of k_2^*, nor is $q_{c_2}(k_1^*,k_2)$ of k_2^*. This implies that each candidate maximizes her utility with respect to her offer. Thus,*

1. $q_{c_1}^* = \sigma_{c_1}$ *if* $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1$
2. $q_{c_2}^* = (B,0)$ *if* $\frac{\upsilon_1}{\upsilon_2} \leq \varepsilon_A \leq 1$

If $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$, then $q_{c_1}(k_1,k_2^*)$ and $q_{c_2}(k_1^*,k_2)$ are indeed functions of the strategy of the other candidate. Holding them as given, and reacting accordingly leads to the following.

1. $q_{c_1}^* = (B,0)$ *if* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$
2. $q_{c_2}^* = \left(B, 1 - \frac{\upsilon_r - \varepsilon_i \upsilon_s}{\upsilon_s - \varepsilon_i \upsilon_r}\right)$ *if* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_2}$.

Steps 2 and 4 lead to Theorem 6. $\square$

11.2 Two term model

We consider a dynamic model where two politicians compete for candidacy in parties, offering a share of their ruling power. Time is discrete, and each period t in $\{1, 2, 3, \ldots\}$, politicians are matched to candidates. Incumbents run for re-election if they have been in office for one term, and they vacate their position after their second term. We proceed with the analysis when an incumbent runs for re-election.

We assume that the incumbent runs for re-election under the governing party's label, and that two politicians compete for the candidacy on the opposition's side. They are myopic, in the sense that they only consider the utility of the current period, and disregard the fact that if they were to win the election, they would run for a second term. The two politicians, $c_{1,t}$ and $c_{2,t}$, denoted by c_1 and c_2, whenever there is no confusion, are characterized by their respective valences υ_1 and υ_2. Valences are random variables that take positive values. The electorate prefers high over low valence candidates. We formalize the per period competition between politicians for the candidacy of the opposition with the following three stage game.

In stage 1, nature determines the valences of politicians c_1 and c_2, υ_1 and υ_2. The realizations are observed by both politicians and the opposition. In stage 2, politicians make simultaneous offers to the opposition. So, the strategy of politician c in $\{c_1, c_2\}$, denoted by q_c, is q_c in $[0, 1]$, where $q_c = k$ means that politician c makes an offer to the opposition, proposing to keep k percent of power if she wins. In stage 3, the opposition picks the offer that maximizes its utility. We denote the strategy of the opposition i in $\{A, B\}$ by s_i in $\{O^i_{c_1}, O^i_{c_2}\}$, where O^i_c means that the unseated party i accepts the offer of politician c. Let σ_i be the mixed strategy of i, that is, $\sigma_c = (p_{c_1}, 1-p_{c_1}$, where $0 \leq p_{c_1} \leq 1$ is the probability that i accepts the offer of politician c_1, and $0 \leq 1 - p_{c_1} \leq 1$ is the probability that it accepts the offer of c_2. Once the candidate and party assignments are made, the payoffs are observed.

We continue to assume that the utilities of the winning politician and her party are proportional to the margin of victory. Thus, they are the same as those given in (46). The extensive-form game is portrayed in Figure 13.

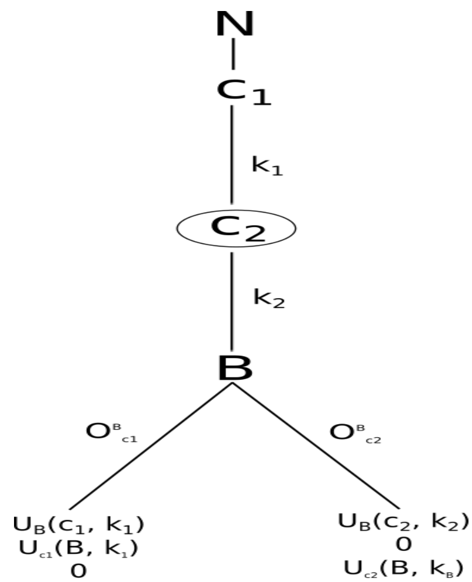

Figure 13: Single term game with offers by candidates

The following result may be proved in much the same way as Theorem 6.

Theorem 7. *Let i in $\{A, B\}$ be the governing party of period t, and let the valence of its incumbent be $\varepsilon_i \upsilon_i$. The requirement on politicians c_r and c_s in $\{c_1, c_2\}$ is that $\upsilon_r < \upsilon_s$.*

1. *If $\frac{\upsilon_r}{\upsilon_i} \leq \varepsilon_i$, then subgame perfect Nash equilibrium is $q^*_{c_r} = 1$, $q^*_{c_s} = 1$, and $s^*_j = O^j_{c_s}$.*

2. *If $\frac{\upsilon_r}{\upsilon_i} > \varepsilon_i$, then subgame perfect Nash equilibria are $q^*_{c_r} = 0$, $q^*_{c_s} = 1 - \frac{\upsilon_r - \varepsilon_i \upsilon_i}{\upsilon_s - \varepsilon_i \upsilon_i}$, and $s^*_j = O^j_{c_s}$.*

Proof. Throughout the proof, we assume that the valences of c_1 and c_2 are such that $\upsilon_1 < \upsilon_2$, and that party A is incumbent at time t. We also assume that the candidate of A has valence $\varepsilon_A \upsilon_A$, where ε_A is the incumbency advantage, and υ_A is her initial valence.

Given the sequential nature of the game, we use backward induction to compute the subgame

Preference	$\succeq_B$		
$\frac{m_1}{m_2}$	$< c$	$= c$	$> c$
Most	c_2	c_1/c_2	c_1
Least	c_1	c_2/c_1	c_2

Table 4: Preference order of B

perfect Nash equilibria. From te definition of utilities given in (46) we have

$$U_B(c_1,k_1) = \begin{cases} 0 & \text{if } \varepsilon_A \geq \frac{\upsilon_1}{\upsilon_A} \\ (1-k_1)(\upsilon_1 - \varepsilon_A \upsilon_A) & \text{if } \varepsilon_A < \frac{\upsilon_1}{\upsilon_A} \end{cases} \quad U_B(c_2,k_2) = \begin{cases} 0 & \text{if } \varepsilon_A \geq \frac{\upsilon_2}{\upsilon_A} \\ (1-k_2)(\upsilon_2 - \varepsilon_A \upsilon_A) & \text{if } \varepsilon_A < \frac{\upsilon_2}{\upsilon_A} \end{cases}. \tag{47}$$

Step 1. *From the above utilities, we have that if $\varepsilon_A \geq \frac{\upsilon_2}{\varepsilon_A}$, then $U_B(c_1,k_1) = U_B(c_2,k_2) = 0$ for all $k_1, k_2 \geq 0$. If $\frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A < \frac{\upsilon_2}{\upsilon_A}$, then $0 = U_B(c_1,k_1) < U_B(c_2,k_2)$ for all $k_1 \geq 0$, and $k_2 > 0$. Lastly, if $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$, then $U_B(c_1,k_1) = (1-k_1)(\upsilon_1 - \varepsilon_A\upsilon_A)$, and $U_B(c_2,k_2) = (1-k_2)(\upsilon_2 - \varepsilon_A\upsilon_A)$.* In other words, if $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$, then the preference of B is determined by the offers it gets from the politicians. Table 2 shows the preference of B when $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$, as a function of the ratio $\frac{m_1}{m_2}$, where $m_1 = 1 - k_1$ is the percentage of power B gets if c_1 wins the election, and $m_2 = 1 - k_2$ is the percentage it gets if c_2 does. We define $c = \frac{\upsilon_2 - \varepsilon_A\upsilon_A}{\upsilon_1 - \varepsilon_A\upsilon_A}$ to be the indifference curve of power offers for B, it is obtained by equalizing $U_B(c_1,k_1)$ and $U_B(c_2,k_2)$, and clearing the ratio $\frac{m_1}{m_2}$. The information is displayed graphically in Figure 14.

The interpretation of Figure 14 leads to the following step.

Step 2. *We define $c := \frac{\upsilon_2 - \varepsilon_A\upsilon_A}{\upsilon_1 - \varepsilon_A\upsilon_A}$.*

- *$s_B(q_{c_1}, q_{c_2}) = \sigma_B$ for all $\varepsilon_A \geq \frac{\upsilon_2}{\upsilon_A}$, and $k_1, k_2 \geq 0$*
- *$s_B(q_{c_1}, q_{c_2}) = O^B_{c_2}$ for all $\frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A < \frac{\upsilon_2}{\upsilon_A}$, $k_1 \geq 0$, and $k_2 > 0$*
- *$s_B(q_{c_1}, q_{c_2}) = O^B_{c_2}$ for all $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$, and $\frac{1-k_1}{1-k_2} < c$*

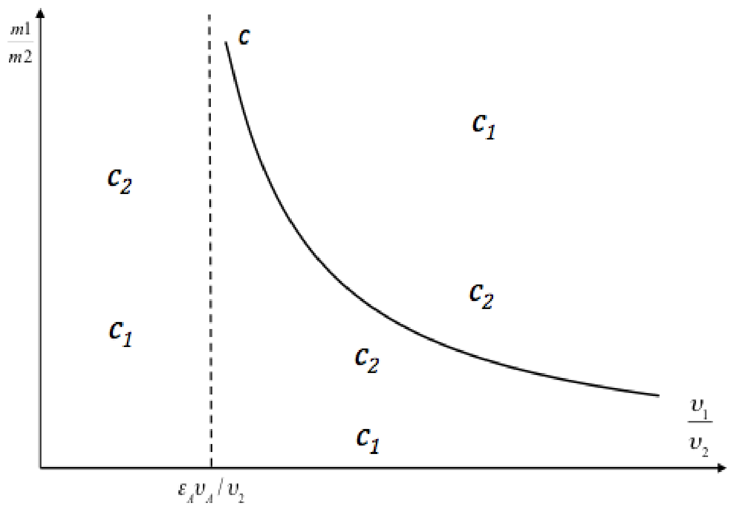

Figure 14: Preferences of B

- $s_B(q_{c_1}, q_{c_2}) = O^B_{c_1}$ *for all* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$, *and* $\frac{1-k_1}{1-k_2} > c$
- $s_B(q_{c_1}, q_{c_2}) = \sigma_B$ *for all* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$, *and* $\frac{1-k_1}{1-k_2} = c$

Given the best responses of the opposition, our next concern is the decision-making of politicians in stage 2. We turn to the definition of utilities given in (44), plugging the dominant strategies of the opposition established at step 2, we find that the only strictly positive utilities of politicians are

$$\begin{aligned}
U_{c_1}((B,k_1) \mid s_B(k_1,k_2)) &= k_1(\upsilon_1 - \varepsilon_A \upsilon_A) \quad \text{if } \varepsilon_A < \frac{\upsilon_1}{\upsilon_A}, \text{ and } \frac{1-k_1}{1-k_2} \geq c \\
U_{c_2}((B,k_2) \mid s_B(k_1,k_2)) &= k_2(\upsilon_2 - \varepsilon_A \upsilon_A) \quad \text{if } \frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A < \frac{\upsilon_2}{\upsilon_A}, \text{ and } \frac{1-k_1}{1-k_2} \geq 0 \\
U_{c_2}((B,k_2) \mid s_B(k_1,k_2)) &= k_2(\upsilon_2 - \varepsilon_A \upsilon_A) \quad \text{if } \varepsilon_A < \frac{\upsilon_1}{\upsilon_A} \text{ and } \frac{1-k_1}{1-k_2} \leq c.
\end{aligned}$$

Note that $0 < U_{c_1}((B,k_1) \mid s_B(k_1,k_2))$ if $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$ and $\frac{1-k_1}{1-k_2} \geq c$, and $0 < U_{c_2}((B,k_2) \mid s_B(k_1,k_2))$ if $\frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A < \frac{\upsilon_2}{\upsilon_A}$ and $\frac{1-k_1}{1-k_2} \geq 0$, and $0 < U_{c_2}((B,k_2) \mid s_B(k_1,k_2))$ if $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$ and $\frac{1-k_1}{1-k_2} \leq c$. This leads to the following strategies of politicians.

Step 3. *Set* $c = \frac{\upsilon_2 - \varepsilon_A \upsilon_A}{\upsilon_1 - \varepsilon_A \upsilon_A}$, *the strategies of politicians* c_1 *and* c_2, q_{c_1} *and* q_{c_2}, *are*

- $q_{c_1}(k_1, k_2) = k_1$ *with* k_1 *in* $[0, 1]$ *if* $\frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A$

- $q_{c_1}(k_1, k_2) = k_1$ *with* k_1 *such that* $\frac{1-k_1}{1-k_2} \geq c$ *if* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$

- $q_{c_2}(k_1, k_2) = k_2$ *with* k_2 *in* $[0, 1]$ *if* $\frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A$

- $q_{c_2}(k_1, k_2) = k_2$ *with* k_2 *such that* $\frac{1-k_1}{1-k_2} \leq c$ *if* $\frac{\upsilon_1}{\upsilon_A} > \varepsilon_A$

Our last step concerns the utility maximization of politicians. We next analyze their best responses. ***Step 4.*** *First note that if* $\frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A$*, then* $q^*_{c_1}(k_1, k^*_2)$ *is not a function of* k^*_2*, nor is* $q^*_{c_2}(k^*_1, k_2)$ *of* k^*_1*. This implies that each candidate maximizes her utility with respect to her own offer. Thus,*

1. $q^*_{c_1} = 1$ *if* $\frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A$

2. $q^*_{c_2} = 1$ *if* $\frac{\upsilon_1}{\upsilon_A} \leq \varepsilon_A$

If $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$*, then* $q^*_{c_1}(k_1, k^*_2)$ *and* $q^*_{c_2}(k^*_1, k_2)$ *are functions of the strategy of the other politician. Holding them as given, and reacting accordingly leads to the following.*

1. $q^*_{c_1} = 0$ *if* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$

2. $q^*_{c_2} = 1 - \frac{\upsilon_1 - \varepsilon_A \upsilon_A}{\upsilon_2 - \varepsilon_A \upsilon_A}$ *if* $\varepsilon_A < \frac{\upsilon_1}{\upsilon_A}$.

Steps 2 and 4 lead to Theorem 7. □

12 Simulations

It is our interest to study the long-run implications of term limits in terms of the average time spent by parties in power, the relative competence of the winner, and the share of power. To this end, we program the dynamics of each model: single term, single term with party offers, single term with candidate offers, two term, two term with party offers, and two term with candidate offers, and run 50 simulations of elections. From these, we obtain measures like the average time that parties spend in power, the percentage of elections in which the lesser candidate is elected, and the power share of parties, and compare them across all the models. The R code for the simulations as well as a brief explanation of the programs can be found in Appendix H.

12.1 Convergence

From chapter 2, we know that if there is a single term limit and voters always prefer alternation, or alternatively, if there is negative partisan advantage in every election, then each party is expected to spend half of the time in office in the long run. We find that this is also the case in the rest of the models, nevertheless, the speed of convergence varies according to the limit of terms.

Figure 15 plots for each dynamics, the per period average number of victories of party A when there is negative partisan advantage. In these simulations, we have assumed that the valences of candidates have exponential distribution of parameter 1, and that the partisan advantages have uniform distribution between 0 and 1.

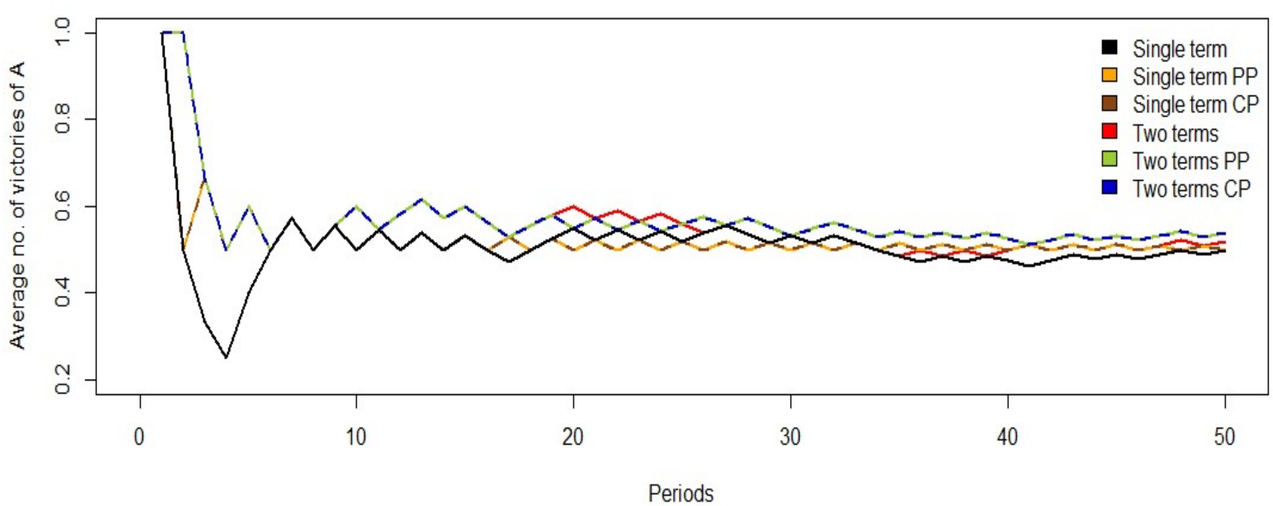

Figure 15: Stationary state with negative advantage

One sees that all three single term models have a faster convergence to 0.5, which is the stationary state obtained in chapter 2, while the three two term models have averages above the analytical prediction. The intuition behind this result is that incumbents that win in spite of being affected by a negative advantage, are positively selected, and therefore perform well against other contenders if they run for a second term in office. This increases the number of victories of the winning party, and explains why the stationary state of two terms is above (or below, in case the winning party is the opposition B) the one of single term.

Another prediction of chapter 2, is that if voters have a high preference for stability, or in other words, if the partisan incumbency advantage is positive and very large, then the first winning party has a high chance of staying in power, and therefore, its stationary state is one. Figure 16 plots the per period average number of victories of A when there is positive partisan advantage. In these simulations, we have assumed that the valences of candidates are exponentially distributed with rate 1, and that the partisan advantages are uniformly distributed between 40 and 50.

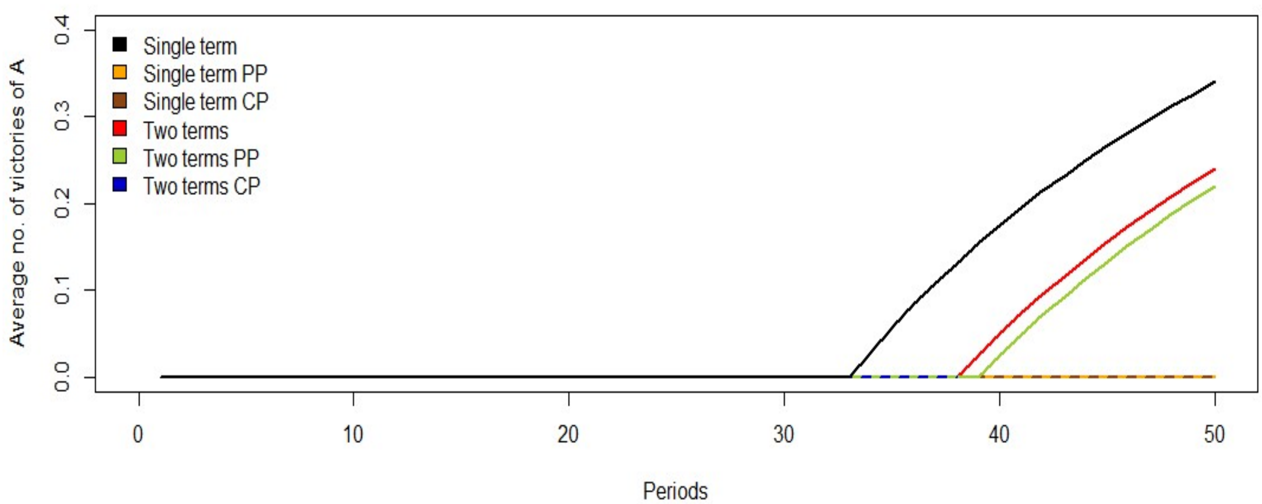

Figure 16: Stationary state with large positive advantage

It can be seen that all 6 models abide by the theoretical prediction of chapter 2. In this case, however, the result needs to be reinterpreted, as the first winning party is not A, but its opponent, B. Given the little preference for alternation of voters, alternation is not likely to occur, as is the case in the first 33 elections. It is worth noting however, that transitions are in fact possible, since they have a very small yet positive probability of occurring. This explains the increase in the average number of victories of A in the right part of the graph.

Lastly, we present the stationary state of A for in-between values of partisan advantage. The conjecture is that if voters are not consecutively in favor of alternation, or against it, then we should expect a more frequent alternation of parties than the absorbing state case, but less than when there is perfect alternation. That is, the average number of victories can range anywhere between 0 and 1, depending on the realizations of the advantages. Figure 17 shows the per period average number of victories of A when there are both positive and negative advantages. We have assumed that the valences of candidates are exponential with rate 1, and that the advantages have uniform distribution between 0 and 2.

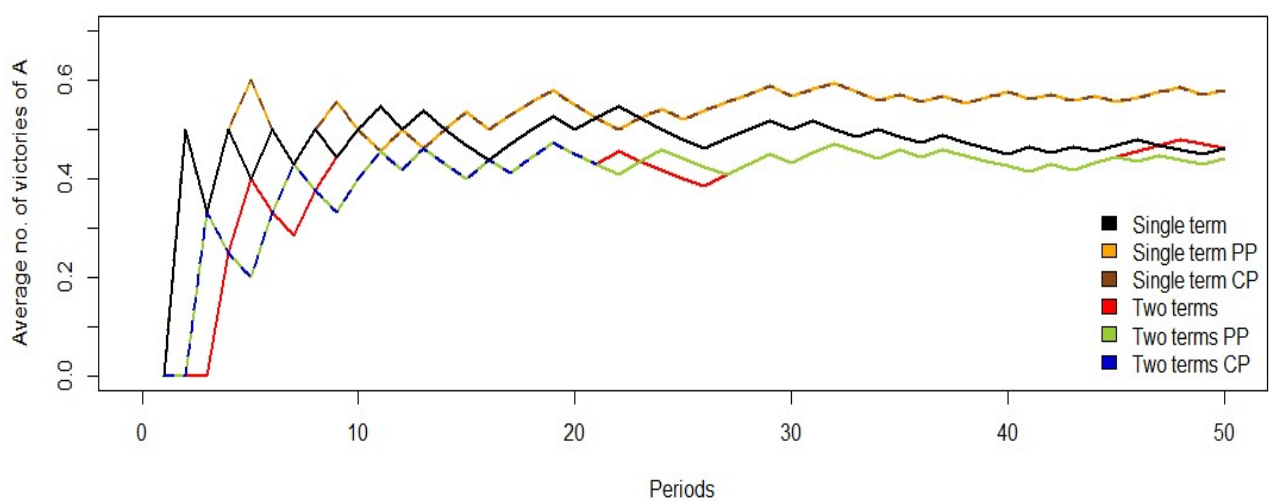

Figure 17: Stationary state with mixed advantages

One sees that the single term models with party and candidate offers have a higher average number of victories than the single term model alone. This is because the single term model is not strategic, in the sense that it does not consider the most beneficial matchings between candidates and parties, whereas the former models do. However, the two term models do not exhibit this monotonic improvement with respect to the single term model. A reason for this is that even though the benefits of running for re-election with a competent candidate are clear, the disadvantages of running with the less competent politician are also straightforward. Explicitly, a party can win an election not because its candidate is the best, but because voters wanted alternation or were in favor of stability. Such cases lead to a politician with low valence running for re-election, possibly against a better opponent than him. This is why the effects of two terms depend on what type of advantage got the incumbent in office.

We next analyze the effects of partisan incumbency advantage on the relative competence of the winner.

12.2 Efficiency

To measure the efficiency of a model, we compute the percentage of elections in which voters choose the candidate with the lowest valence. Since we are interested in the effect of partisan advantage, we sample advantages from different distributions and obtain the measure of efficiency of each. We have assumed that valences are drawn from an exponential distribution of rate 1, and that the advantages are drawn from uniform distributions ranging from (0,1) to (0,20). Figure 18 shows the percentage of elections in which the lesser candidate is chosen.

Note that the single term models with party and candidate offers have the lowest percentage of elections choosing the lowest valence valued candidate, or alternatively, have the highest efficiency. This is consistent with the observation that agents that make the offers chose the option that grants them the highest chance of winning, so the assignments that

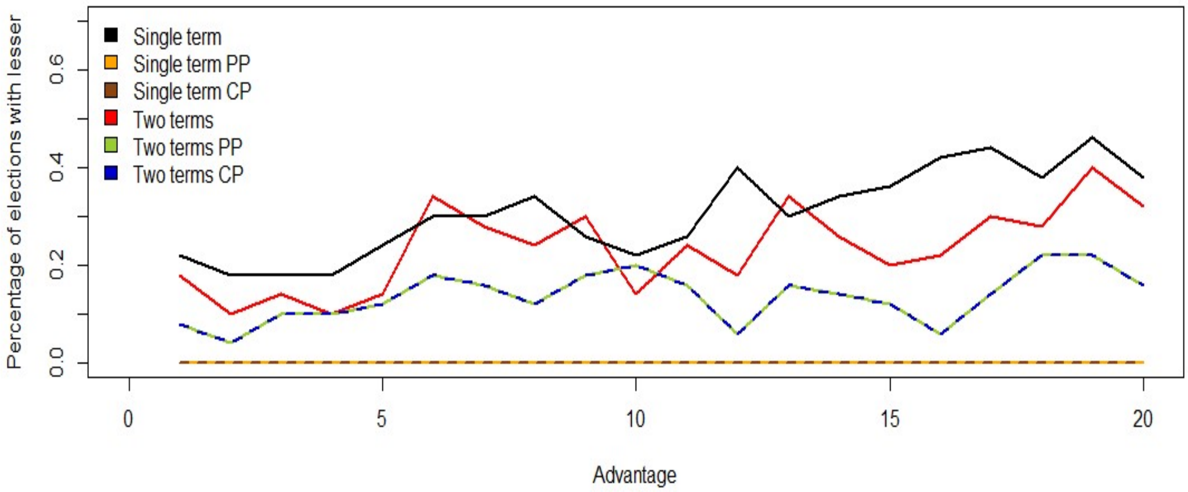

Figure 18: Efficiency with varying advantages

result from bargaining are in fact optimal. Recall that in the single term model with offer making by parties, the highest valued candidate is matched to the advantaged party, whether it be the ruling party if there is positive partisan advantage, or the opposition if the advantage is negative. In the model with offer making by candidates, the advantaged party is also matched to the most competent candidate, the only difference being the shares of power agreed to. Since efficiency accounts only for the matches themselves and not the agreement under which they took place, the measure of both models is the same.

Notably, this is not observed in either of the two term models with offer-making. The reason is that all agents are myopic, meaning that they only care about the current period's utilities. Consequently, they do not take into account the re-election scenario, and may end up running with a less competitive candidate for the second period. In addition, whenever there is a re-election, the opposition is virtually forced to accept its contender, and therefore is at a relative disadvantage with respect to the sitting party. Recall that in the two term model with offer-making by parties, the opposition chooses between two candidates. If they are both low-valued, at least with respect to the incumbent, the opposition is left to pick the lesser of two evils. It is still the case however, that despite the described inefficiencies, the percentage of elections in which the less competitive candidate is chosen is below that of the single term and the two term limits.

It is also worth pointing out that for the single and two term models, efficiency decreases as the partisan advantage increases. The idea is that higher advantages translate into voters being more prone to stability, so they are more likely to vote for the incumbent or the successor for the sake of being loyal to the sitting party rather than for merit. This logic also explains the relative peaks to the left of the graphs, where voters are partial to alternation, and vote for challengers for the sake of trying something new. The lowest percentages are those closer to an advantage of 1, or to unbiased voting.

12.3 Power share

Our last concern is the amount of power of agents as a function of candidate valence. One may conjecture that increasing the valence parameter leads to a higher bargaining power of candidates (on average) and thus, to a higher share of power, at least in the models where parties make offers. Nevertheless, since the candidate valences are drawn from the same distribution, one could also expect that the distance between the realizations is relatively preserved. A good indicator of the effect of the increase of the parameter in the distance between observations is variance, and since we have assumed that the distribution of the valences is exponential, we know that variance decreases in the parameter.[3] This implies that as we increase the average candidate valence, the power share function becomes flatter. Conversely, if we decrease it, the function is expected to have more variation. Figure 19 confirms this logic. We have ranged the valence parameter from 0.1 to 2, that is, we have increased exponentially and also halved the average valence of candidates with respect to the reference value which was 1, and we have plotted the average power of parties.

[3]Recall that if X has exponential distribution with rate λ, then $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

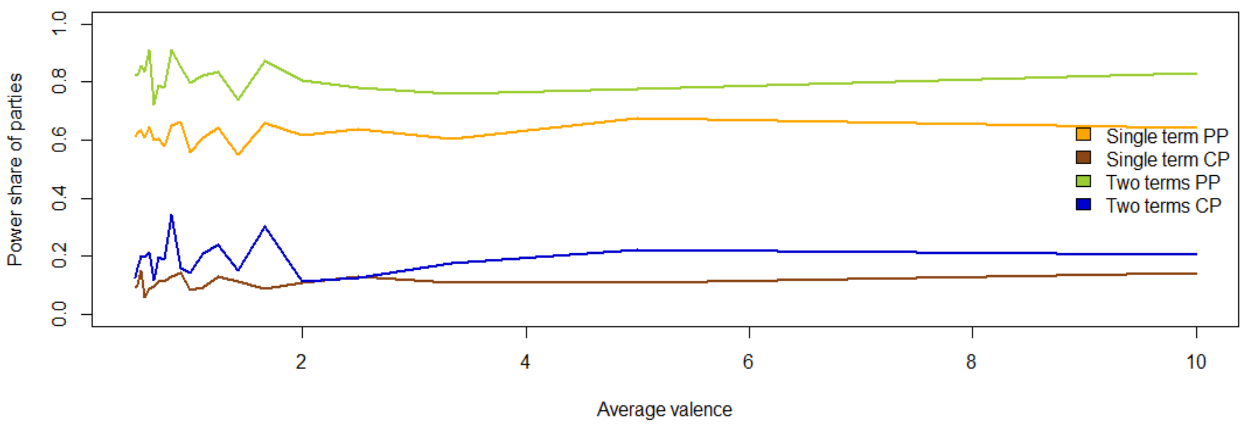

Figure 19: Share of power

One sees that parties get a bigger share of power when they propose to candidates than when they are proposed to, and this holds for all the valence values displayed. The general idea is that parties take advantage of the preference of candidates and offer as little as possible to secure their preferred match. Candidates do the same and maximize their utility by lowering the share of parties just enough so that they still prefer to sign with them. We conclude that in our election games, suitors have a bigger impact on the power shares than term limits do. In other words, setting the political agenda has a larger effect on the empowering of candidates than the possibility of reelection. In this sense, the 2014 electoral reform is expected to have a minor impact on the political life in Mexico.

We now analyze the differences between single term and two terms when parties propose, and single term and two terms when candidates propose. We start by noting that in both scenarios, the power share of parties is higher when there are two terms than when there is a single term limit. Since the shares are the same in every election with single term and an open election in the two term, we focus on the case when an incumbent runs for office.

In the parties propose dynamics there are two possibilities: either the ruling party wins, in which case, it matches the previous offer, or the opposition wins and keeps all the power to itself. This last possibility implies that if the incumbent were to seek and win re-election, then the ruling party would keep all the power for a second time. So, not only does the opposition benefit from monopolistic power in re-elections, but manages to keep it in case of a second victory. On the other hand, in the single term model, the preferred party has to grant its candidate either her indifference share, or no share in the best case scenario, but does this every single period. In other words, competition is inevitable in the single term model, and this drives the share of parties down.

In candidate offer-making models, competition is observed in the two term limit case, meaning that in re-elections, the highest valued candidate does not enjoy monopolistic power. Indeed, her bargaining power is lowered if the second best candidate stands a chance of winning. In this case, she agrees to receive the indifference share. Nevertheless, the possibility of matching a previous share of power through re-election is still present, this potentially explains the empirical result.

References

[1] Aldrich, John H. 1995. *Why parties? The origin and transformation of political parties in America*. University of Chicago Press.

[2] Bawn, Kathleen, Martin Cohen, David Karol, Seth Masket, Hans Noel, and John Zaller. 2012. A theory of political parties: groups, policy demands, and nominations in American politics. *Perspectives on Politics* 10, no. 3: 571-597.

[3] Carey, John M., Richard G. Niemi, Lynda W. Powell, and Gary F. Moncrief. 2006. The effects of term limits on State Legislatures: a new survey of the 50 States. *Legislative Studies Quarterly* 31, no. 1: 105-134.

[4] Cox, Gary, and Mathew D. McCubbins. 1993. *Legislative Leviathan: party government in the House*. University of California Press.

[5] Desmarais, Bruce A., Raymond J. La Raja, and Michael S. Kowal. 2015. The Fates of challengers in U.S. House Elections: the role of extended party networks in supporting candidates and shaping electoral outcomes. *American Journal of Political Science* 59, no. 1: 194-211.

[6] Downs, Anthony. 1957. An Economic Theory of Political Action in a Democracy. *Journal of Political Economy* 65: 135-150.

[7] Glazer A., and Wattenberg M.P. 1996. How will term limits affect legislative work? In *Legislative Term Limits: Public Choice Perspectives* Grofman B. (eds). Studies in Public Choice, vol 10. Springer, Dordrecht

[8] Flowers, Julianne F., Audrey A. Haynes, and Michael H. Crespin. The media, the campaign, and the message. *American Journal of Political Science* 47, no. 2: 259-273.

[9] Green-Pedersen, Christoffer, and Peter B. Mortensen. 2010. Who sets the agenda and who responds to it in the Danish Parliament? A new model of issue competition and agenda-setting. *European Journal of Political Research* 49, no. 2: 257–81.

[10] Holcombe, Randall G., and James D. Gwartney. Political parties and the legislative principal-agent relationship. *Journal of Institutional and Theoretical Economics* 145, no. 4: 669-675.

[11] Kahn, Kim F. 1994. The distorted mirror: press coverage of women candidates for Statewide Office. *The Journal of Politics* 56, no. 1: 154-173.

[12] Klüver, Heike, and Iñaki Sagarzazu. 2016. Setting the agenda or responding to voters? Political parties, voters and issue attention. *West European Politics* 39, no. 2: 380–98.

[13] McCombs, Maxwell. 2004. *Setting the agenda: the mass media and public opinion.* Blackwell Publishers.

[14] Serra, Gilles. 2010. Polarization of What? A Model of Elections with Endogenous Valence. *The Journal of Politics* 72, no. 2: 426-437.

[15] Serra, Gilles. 2011. Why primaries? The party's tradeoff between policy and valence. *Journal of Theoretical Politics* 23, no. 1: 21-51.

[16] Serra, Gilles. 2018. The electoral strategies of a populist candidate: Does charisma discourage experience and encourage extremism? *Journal of Theoretical Politics* 30, no. 1: 45-73.

[17] Stokes, Donald E. 1963. Spatial Models of Party Competition. *The American Political Science Review* 57, no. 2: 368-377.

[18] Skinner, Richard M., Seth E. Masket, and David A. Dulio. 2012. 527 committees and the Political Party Network. *American Politics Research* 40, no. 1: 60-84.

Appendix F Single term model

Offer making by parties, positive advantage case

Theorem 8. *Let i in $\{A, B\}$ be the incumbent party of period t, and let $1 < \varepsilon_i$ be its incumbency advantage. The requirement on candidates c_r and c_s in $\{c_1, c_2\}$ is that $\upsilon_r < \upsilon_s$.*

1. *If $\varepsilon_i \leq \frac{\upsilon_s}{\upsilon_r}$, then subgame perfect Nash equilibrium is unique and is $s_i^* = \left(c_s, \frac{\upsilon_s - \varepsilon_i \upsilon_r}{\varepsilon_i \upsilon_s - \upsilon_r}\right)$, $s_j^* = (c_s, 1)$, $q_{c_r}^* = \sigma_{c_r}$, and $q_{c_s}^* = O_i^{c_s}$.*

2. *If $\frac{\upsilon_s}{\upsilon_r} < \varepsilon_i$, then subgame perfect Nash equilibria are $s_i^* = (c_s, 0)$, $s_j^* = \sigma_j$ with k in $[0, 1]$, $q_{c_r}^* = \sigma_{c_r}$, and $q_{c_s}^* = O_i^{c_s}$.*

Preference	$\succeq_{c_2}$		
$\frac{k_A}{k_B}$	$< c$	$= c$	$> c$
Most	B	A/B	A
Least	A	B/A	B

Table 5: Preference order with positive incumbency advantage

Proof. Throughout the proof, we assume that the valences of c_1 and c_2 are such that $\upsilon_1 < \upsilon_2$, and that party A is incumbent at time t. This involves no loss of generality. The utilities given in (42) and (43) give rise to the following observation.

Step 1. *If* $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$*, then* $U_{c_1}(A, k_A) = U_{c_1}(B, k_B) = 0$ *for all* $k_A, k_B \geq 0$*. If* $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$*, then* $U_{c_1}(B, k_B) < U_{c_1}(A, k_A)$ *for all* $k_A > 0$*, and* $k_B \geq 0$*. If* $\varepsilon_A \geq \frac{\upsilon_2}{\upsilon_1}$*, then* $U_{c_2}(B, k_B) < U_{c_2}(A, k_A)$ *for all* $k_A > 0$*, and* $k_B \geq 0$*. Lastly, if* $1 < \varepsilon_A < \frac{\upsilon_2}{\upsilon_1}$*, then* $U_{c_2}(A, k_A) = k_A(\varepsilon_A \upsilon_2 - \upsilon_1)$*, and* $U_{c_2}(B, k_B) = k_B(\upsilon_2 - \varepsilon_A \upsilon_1)$*.*

Table 5 shows the preference of c_2 when $1 < \varepsilon_A < \frac{\upsilon_2}{\upsilon_1}$, as a function of the ratio of offers $\frac{k_A}{k_B}$. We define $c = \frac{\upsilon_2 - \varepsilon_A \upsilon_1}{\varepsilon_A \upsilon_2 - \upsilon_1}$ to be the indifference curve of offers, obtained by equalizing $U_{c_2}(A, k_A)$ and $U_{c_2}(B, k_B)$, and clearing the ratio $\frac{k_A}{k_B}$. The information is displayed graphically in Figure 20. From these preference we obtain the following strategies of candidates.

Step 2. *We define* $c := \frac{\upsilon_2 - \varepsilon_A \upsilon_1}{\varepsilon_A \upsilon_2 - \upsilon_1}$*. The strategies of* c_1 *and* c_2*,* q_{c_1} *and* q_{c_2}*, are*

- $q_{c_1}(s_A, s_B) = \sigma_{c_1}$ for all $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$ and $\frac{k_A}{k_B} \geq 0$
- $q_{c_1}(s_A, s_B) = O_A^{c_1}$ for all $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$ and $\frac{k_A}{k_B} > 0$
- $q_{c_2}(s_A, s_B) = O_B^{c_2}$ for all $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$ and $\frac{k_A}{k_B} < c$
- $q_{c_2}(s_A, s_B) = O_A^{c_2}$ for all $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$ and $\frac{k_A}{k_B} > c$
- $q_{c_2}(s_A, s_B) = \sigma_{c_2}$ for all $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$ and $\frac{k_A}{k_B} = c$
- $q_{c_2}(s_A, s_B) = O_A^{c_2}$ for all $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$ and $\frac{k_A}{k_B} > 0$.

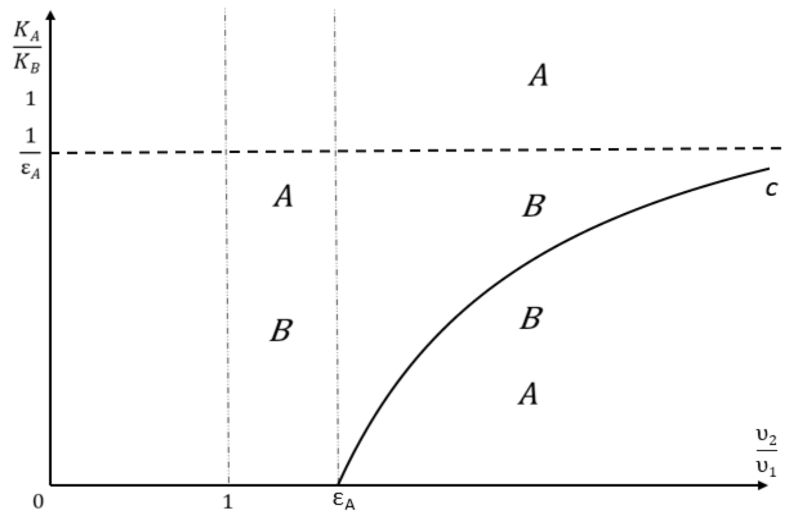

Figure 20: Preferences of c_2 with positive incumbency advantage

Given the strategies of candidates, we analyze the decision-making of parties in stage 2. Here again, we turn to the definition of utilities given in (41), plugging the dominant strategies of candidates established at step 2, we find that the only strictly positive utilities are

$$
\begin{aligned}
U_A((c_1,k_A) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) &= (1-k_A)(\varepsilon_A\upsilon_1-\upsilon_2) \quad \varepsilon_A > \frac{\upsilon_2}{\upsilon_1} \text{ and } \frac{k_A}{k_B} > 0 \\
U_A((c_2,k_A) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) &= (1-k_A)(\varepsilon_A\upsilon_2-\upsilon_1) \quad \text{if } 1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1} \text{ and } \frac{k_A}{k_B} \geq c \\
U_A((c_2,k_A) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) &= (1-k_A)(\varepsilon_A\upsilon_2-\upsilon_1) \quad \text{if } \varepsilon_A > \frac{\upsilon_2}{\upsilon_1} \text{ and } \frac{k_A}{k_B} > 0 \\
U_B((c_2,k_B) \mid q_{c_1}(k_A,k_B), q_{c_2}(k_A,k_B)) &= (1-k_B)(\upsilon_2-\varepsilon_A\upsilon_1) \quad \text{if } 1 < \varepsilon_A < \frac{\upsilon_2}{\upsilon_1} \text{ and } \frac{k_A}{k_B} \leq c.
\end{aligned}
$$

This leads to the following strategies of parties.

Step 3. *Let* $c = \frac{\upsilon_2-\varepsilon_A\upsilon_1}{\varepsilon_A\upsilon_2-\upsilon_1}$. *If* $\varepsilon_A > 1$, *the strategies of parties A and B,* s_A *and* s_B, *are*

- $s_A(k_A,k_B) = (c_2,k_A)$ with k_A such that $k_A \leq ck_B$ if $\varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$
- $s_A(k_A,k_B) = \sigma_A$ if $\frac{\upsilon_2}{\upsilon_1} < \varepsilon_A$
- $s_B(k_A,k_B) = (c_2,k_B)$ with k_B such that $ck_B \leq k_A$ if $\varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$

- $s_B(k_A, k_B) = (c_2, k_B)$ with k_B in $(0, 1]$ if $\frac{\upsilon_2}{\upsilon_1} < \varepsilon_A$.

We proceed with the analysis of best responses.

Step 4. *First note that if $\frac{\upsilon_2}{\upsilon_1} < \varepsilon_A$, then $s_A(k_A, k_B^*)$ is not a function of k_B^*, nor is $s_B(k_A, k_B^*)$ of k_A^*. Thus,*

1. *$s_A^*(k_A, k_B) = \sigma_A$ if $\frac{\upsilon_2}{\upsilon_1} < \varepsilon_A$*

2. *$s_B^*(k_A, k_B) = (c_2, 0)$ if $\frac{\upsilon_2}{\upsilon_1} < \varepsilon_A$.*

If $\frac{\upsilon_2}{\upsilon_1} \leq \varepsilon_A$, then $s_A(k_A, k_B^*)$ and $s_B(k_A, k_B^*)$ are functions of the strategy of the other party. This implies

1. $s_A^*(k_A, k_B) = \left(c_2, \frac{\upsilon_2 - \varepsilon_A \upsilon_1}{\varepsilon_A \upsilon_2 - \upsilon_1}\right)$ if $\frac{\upsilon_2}{\upsilon_1} \leq \varepsilon_A$

2. $s_B^*(k_A, k_B) = (c_2, 1)$ if $\frac{\upsilon_2}{\upsilon_1} \leq \varepsilon_A$.

Steps 2 and 4 yield Theorem 8. □

Appendix G Single term model

Offer making by candidates, positive advantage case

Theorem 9. *Let i in $\{A, B\}$ be the incumbent party of period t, and let $\varepsilon_i > 1$ be its incumbency advantage. The requirement on candidates c_r and c_s in $\{c_1, c_2\}$ is that $\upsilon_r < \upsilon_s$.*

1. *If $\varepsilon_i \leq \frac{\upsilon_s}{\upsilon_r}$, then the subgame perfect Nash equilibria are $s_i^* = O_{c_s}^i$, $s_j^* = \sigma_j$, $q_{c_r}^* = \sigma_{c_r}$, and $q_{c_s}^* = (i, 1)$.*

2. *If $\frac{\upsilon_s}{\upsilon_r} < \varepsilon_i$, then the subgame perfect Nash equilibrium is $s_i^* = O_{c_s}^i$, $s_j^* = \sigma_j$, $q_{c_r}^* = (i, 0)$, and $q_{c_s}^* = \left(i, 1 - \frac{\varepsilon_i \upsilon_r - \upsilon_s}{\varepsilon_i \upsilon_s - \upsilon_r}\right)$.*

Preference	$\succeq_A$		
$\frac{m_1}{m_2}$	$< c$	$= c$	$> c$
Most	c_2	c_1/c_2	c_1
Least	c_1	c_2/c_1	c_2

Table 6: Preference order with positive incumbency advantage

Proof. We continue to assume that the valences of c_1 and c_2 are such that $\upsilon_1 < \upsilon_2$, and that party A is incumbent at time t. The utilities given in (45) and (46) give rise to the following observation.

Step 1. *If $1 \leq \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$, then $U_A(c_1, k_1) = 0 < U_A(c_2, k_2)$ for all $k_1 \geq 0$ and $k_2 > 0$. If $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$, then $U_A(c_1, k_1) = (1 - k_1)(\varepsilon_A \upsilon_1 - \upsilon_2)$, and $U_A(c_2, k_2) = (1 - k_2)(\varepsilon_A \upsilon_2 - \upsilon_1)$. If $1 \leq \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$, then $U_B(c_1, k_1) = 0 < U_B(c_2, k_2)$ for all $k_1 \geq 0$ and $k_2 > 0$. If $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$, then $U_B(c_1, k_1) = U_B(c_2, k_2) = 0$ for all $k_1, k_1 \geq 0$.*

Table 6 shows the preference of A when $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$, as a function of the ratio of offers $\frac{m_1}{m_2} = \frac{1-k_1}{1-k_2}$. We define $c = \frac{\varepsilon_A \upsilon_2 - \upsilon_1}{\varepsilon_A \upsilon_1 - \upsilon_2}$ to be the indifference curve of offers, it is obtained by equalizing $U_A(c_1, k_1)$ and $U_A(c_2, k_2)$, and clearing the ratio $\frac{1-k_1}{1-k_2}$. The information is displayed graphically in Figure 21. From these preferences we obtain the following strategies of parties.

Step 2. *We define $c := \frac{\varepsilon_A \upsilon_2 - \upsilon_1}{\varepsilon_A \upsilon_1 - \upsilon_2}$. The strategies of A and B, s_A and s_B, are*

- $s_A(q_{c_1}, q_{c_2}) = O^A_{c_2}$ *for all* $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$ *and* $\frac{1-k_1}{1-k_2} \geq 0$
- $s_A(q_{c_1}, q_{c_2}) = O^A_{c_2}$ *for all* $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$ *and* $\frac{1-k_1}{1-k_2} < c$
- $s_A(q_{c_1}, q_{c_2}) = O^A_{c_1}$ *for all* $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$ *and* $\frac{1-k_1}{1-k_2} > c$
- $s_A(q_{c_1}, q_{c_2}) = \sigma_A$ *for all* $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$ *and* $\frac{1-k_1}{1-k_2} = c$
- $s_B(q_{c_1}, q_{c_2}) = O^B_{c_2}$ *for all* $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$ *and* $\frac{1-k_1}{1-k_2} \geq 0$
- $s_B(q_{c_1}, q_{c_2}) = \sigma_B$ *for all* $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$ *and* $\frac{1-k_1}{1-k_2} \geq 0.$

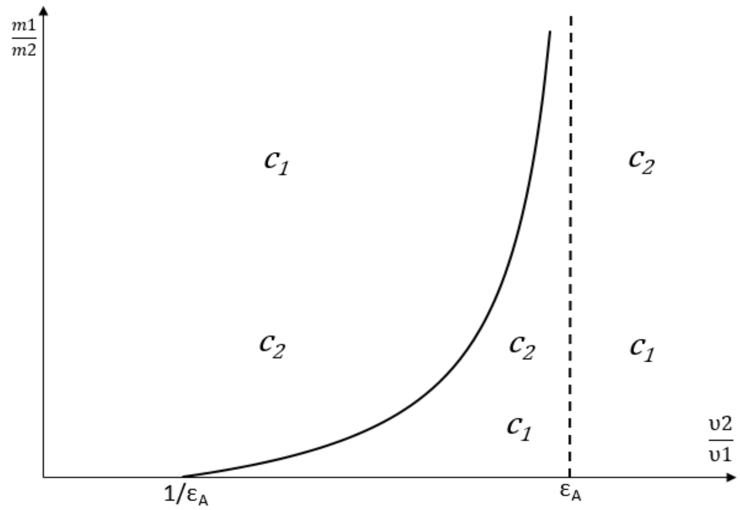

Figure 21: Preferences of A with positive incumbency advantage

Given the best responses of parties, our next concern is the decision-making of candidates in stage 2. We turn to the definition of utilities given in (44), plugging the dominant strategies of parties established at step 2, we find that the only strictly positive utilities are

$$\begin{aligned}
U_{c_1}((A,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2)) &= k_1(\varepsilon_A\upsilon_1 - \upsilon_2) \quad \text{if } \varepsilon_A > \frac{\upsilon_2}{\upsilon_1} \text{ and } \frac{1-k_1}{1-k_2} \geq c \\
U_{c_2}((A,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2)) &= k_2(\varepsilon_A\upsilon_2 - \upsilon_1) \quad \text{if } 1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1} \text{ and } \frac{1-k_1}{1-k_2} \geq 0 \\
U_{c_2}((A,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2)) &= k_2(\varepsilon_A\upsilon_2 - \upsilon_1) \quad \text{if } \varepsilon_A > \frac{\upsilon_2}{\upsilon_1} \text{ and } \frac{1-k_1}{1-k_2} \leq c \\
U_{c_2}((B,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2)) &= k_2(\upsilon_2 - \varepsilon_A\upsilon_1) \quad \text{if } 1 < \varepsilon_A < \frac{\upsilon_2}{\upsilon_1} \text{ and } \frac{1-k_1}{1-k_2} \geq 0.
\end{aligned}$$

Note that $U_{c_1}((B,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2)) < U_{c_1}((A,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2))$ if $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$ and $\frac{1-k_1}{1-k_2} \geq c$, and $U_{c_2}((B,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2)) < U_{c_2}((A,k_1) \mid s_A(k_1,k_2), s_B(k_1,k_2))$ if $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$ and $\frac{1-k_1}{1-k_2} \geq 0$, or if $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$ and $\frac{1-k_1}{1-k_2} \leq c$. This leads to the following strategies of candidates.

Step 3. *Let $c = \frac{\varepsilon_A\upsilon_2 - \upsilon_1}{\varepsilon_A\upsilon_1 - \upsilon_2}$, the strategies of candidates c_1 and c_2, q_{c_1} and q_{c_2}, are*

- $q_{c_1}(k_1,k_2) = \sigma_{c_1}$ *with* $k_1 \geq 0$ *if* $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$

- $q_{c_1}(k_1, k_2) = (A, k_1)$ *with* k_1 *such that* $1 - k_1 \geq c(1 - k_2)$ *if* $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$
- $q_{c_2}(k_1, k_2) = (A, k_2)$ *with* k_2 *in* $(0, 1]$ *if* $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$
- $q_{c_2}(k_1, k_2) = (A, k_2)$ *with* k_2 *such that* $1 - k_1 \leq c(1 - k_2)$ *if* $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$

We proceed with the analysis of best responses.

Step 4. *First note that if* $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$*, then* $q_{c_1}(k_1, k_2^*)$ *is not a function of* k_2^**, nor is* $q_{c_2}(k_1^*, k_2)$ *of* k_1^**. Thus,*

1. $q_{c_1}^* = \sigma_{c_1}$ *if* $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$
2. $q_{c_2}^* = (A, 0)$ *if* $1 < \varepsilon_A \leq \frac{\upsilon_2}{\upsilon_1}$

If $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$*, then* $q_{c_1}(k_1, k_2^*)$ *and* $q_{c_2}(k_1^*, k_2)$ *are indeed functions of the strategy of the other candidate. Holding them as given, and reacting accordingly leads to the following.*

1. $q_{c_1}^* = (A, 1)$ *if* $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$
2. $q_{c_2}^* = \left(A, 1 - \frac{\varepsilon_i \upsilon_r - \upsilon_s}{\varepsilon_i \upsilon_s - \upsilon_r}\right)$ *if* $\varepsilon_A > \frac{\upsilon_2}{\upsilon_1}$

Steps 2 and 4 lead to Theorem 9. □

Appendix H Simulations

In this section the reader will find the R code for simulations of elections. The main idea is to define the distributions of candidate valence and partisan advantage, and to obtain realizations of each in every period: one for each candidate if there is an open seat election, or one for the contender if two terms are permitted and the incumbent is running for reelection, and one for the advantage. We next use the values to determine the winning party, as well as the power share if offers are made by either the parties or the candidates. For this purpose, we Lastly, we use the track of victories of parties to compute the average time spent by each party in power.

For the simulations we assume that candidate valences are exponentially distributed with rate 1, and we also assume that the partisan advantages are uniformly distributed between 0 and 1.

```
# R is the number of simulations of r. v.
# R/2 is the number of periods
R=100
# E is vector of advantages
E=NULL
# X is the vector of candidate valences
X=NULL
# W's are the per period indicators of
# victories of A: 0 if it wins, 0 if not
# W1 Single term, W2 Two terms,
# W3 Single term Parties Propose
# W4 Single term Candidates Propose
# W5 Two terms Parties Propose
# W6 Two terms Candidates propose
W1=NULL
W2=NULL
W3=NULL
W4=NULL
W5=NULL
W6=NULL
# V vector of highest valence per period
V=NULL
# V's vectors of valences of winners
# according to model: 1 to 6
V1=NULL
V2=NULL
V3=NULL
V4=NULL
V5=NULL
V6=NULL
# C's smallest valence ratio
C3=NULL
C4=NULL
C5=NULL
C6=NULL
# G's biggest valence ratio
G3=NULL
G4=NULL
G5=NULL
G6=NULL
# A's power share
A3=NULL
A4=NULL
A5=NULL
A6=NULL
# B's power share
B3=NULL
B4=NULL
B5=NULL
B6=NULL
# S's power share of the winning party
S3=NULL
S4=NULL
S5=NULL
S6=NULL
# P's partial number of
# victories of A
P1=NULL
P2=NULL
P3=NULL
P4=NULL
P5=NULL
P6=NULL
# K's partial average number
# of victories of A
K1=NULL
K2=NULL
K3=NULL
```

```
K4=NULL
K5=NULL
K6=NULL
# N's vectors of number of terms in
power
N2=NULL
N5=NULL
N6=NULL
# D's difference between actual highest
valence and that of winning candidate
D1=NULL
# Obtain R realization of exponential
# random variables of rate 1
# these are candidate valences
for (i in 1:R){
X[i]=rexp(1,rate=1)
}
# Obtain R/2 uniform advantages,
# one per period
k=R/2
for (i in 1:k)
{
E[i]=runif(1, min=0, max=1)
}
# Save highest valence per period
for (i in 1:k){
j=2*i
if (X[j-1]<X[j]){
V[i]=X[j]
} else {
V[i]=X[j-1]
}
}
# First winning party, no advantage
if (X[1]<X[2]){
W1[1]=0
W2[1]=0
W3[1]=0
W4[1]=0
W5[1]=0
W6[1]=0
V1[1]=X[2]
V2[1]=X[2]
V3[1]=X[2]
V4[1]=X[2]
V5[1]=X[2]
V6[1]=X[2]
} else {
W1[1]=1
W2[1]=1
W3[1]=1
W4[1]=1
W5[1]=1
W6[1]=1
V1[1]=X[1]
V2[1]=X[1]
V3[1]=X[1]
V4[1]=X[1]
V5[1]=X[1]
V6[1]=X[1]
}
# Single term begins
for (i in 2:k){
j=2*i
if (W1[i-1]==1){
if (E[i]*X[j-1]>X[j]){
W1[i]=1
V1[i]=X[j-1]
} else if (E[i]*X[j-1]<X[j]){
W1[i]=0
V1[i]=X[j]
}
} else if (W1[i-1]==0){
if (E[i]*X[j]>X[j-1]){
W1[i]=0
V1[i]=X[j]
} else if (E[i]*X[j]<X[j-1]){
W1[i]=1
V1[i]=X[j-1]
}
}
```

```
}
# Single term ends
# Two term begins
N2[1]=1
for (i in 2:k){
j=2*i
if (N2[i-1]==2){
if (W2[i-1]==1){
if (E[i]*X[j-1]>X[j]){
W2[i]=1
V2[i]=X[j-1]
N2[i]=1
} else if (E[i]*X[j-1]<X[j]){
W2[i]=0
V2[i]=X[j]
N2[i]=1
}
} else if (W2[i-1]==0){
if (X[j-1]<E[i]*X[j]){
W2[i]=0
V2[i]=X[j]
N2[i]=1
} else if (X[j-1]>E[i]*X[j]){
W2[i]=1
V2[i]=X[j-1]
N2[i]=1
}
}
} else if (N2[i-1]==1){
if (W2[i-1]==1){
if (E[i]*X[j-3]>X[j]){
W2[i]=1
V2[i]=X[j-3]
N2[i]=N2[i-1]+1
} else if (E[i]*X[j-3]<X[j]){
W2[i]=0
V2[i]=X[j]
N2[i]=1
}
} else if (W2[i-1]==0){
if (X[j-1]<E[i]*X[j-2]){
W2[i]=0
V2[i]=X[j-2]
N2[i]=N2[i-1]+1
} else if (X[j-1]>E[i]*X[j-2]){
W2[i]=1
V2[i]=X[j-1]
N2[i]=1
}
}
}
}
# Two term ends
# Single term parties propose begins
for (i in 2:k){
j=2*i
if (X[j-1]<X[j]){
C3[i]=X[j-1]/X[j]
G3[i]=1/C3[i]
V3[i]=X[j]
} else if (X[j-1]>X[j]){
C3[i]=X[j]/X[j-1]
G3[i]=1/C3[i]
V3[i]=X[j-1]
}
if (W3[i-1]==1){
if (C3[i]<E[i] & E[i]<1){
W3[i]=0
B3[i]=1-(E[i]-C3[i])/(1-E[i]*C3[i])
S3[i]=B3[i]
} else if (E[i]<C3[i]){
W3[i]=0
B3[i]=1
S3[i]=B3[i]
} else if (G3[i]<E[i]){
W3[i]=1
A3[i]=1
S3[i]=A3[i]
} else if (1<E[i] & E[i]<G3[i]){
W3[i]=1
A3[i]=1-(G3[i]-E[i])/(E[i]*G3[i]-1)
```

```
S3[i]=A3[i]
}
} else if (W3[i-1]==0){
if (C3[i]<E[i] & E[i]<1){
W3[i]=1
A3[i]=1-(E[i]-C3[i])/(1-E[i]*C3[i])
S3[i]=A3[i]
} else if (E[i]<C3[i]){
W3[i]=1
A3[i]=1
S3[i]=A3[i]
} else if (G3[i]<E[i]){
W3[i]=0
B3[i]=1
S3[i]=B3[i]
} else if (1<E[i] & E[i]<G3[i]){
W3[i]=0
B3[i]=1-(G3[i]-E[i])/(E[i]*G3[i]-1)
S3[i]=B3[i]
}
}
}
# Single term parties propose ends
# Single term candidates propose begins
for (i in 2:k){
j=2*i
if (X[j-1]<X[j]){
C4[i]=X[j-1]/X[j]
G4[i]=1/C4[i]
V4[i]=X[j]
} else if (X[j-1]>X[j]){
C4[i]=X[j]/X[j-1]
G4[i]=1/C4[i]
V4[i]=X[j-1]
}
if (W4[i-1]==1){
if (C4[i]<E[i] & E[i]<1){
W4[i]=0
B4[i]=0
S4[i]=B4[i]
} else if (E[i]<C4[i]){
W4[i]=0
B4[i]=(C4[i]-E[i])/(1-E[i]*C4[i])
S4[i]=B4[i]
} else if (G4[i]<E[i]){
W4[i]=1
A4[i]=(E[i]-G4[i])/(E[i]*G4[i]-1)
S4[i]=A4[i]
} else if (1<E[i] & E[i]<G4[i]){
W4[i]=1
A4[i]=0
S4[i]=A4[i]
}
} else if (W4[i-1]==0){
if (C4[i]<E[i] & E[i]<1){
W4[i]=1
A4[i]=0
S4[i]=A4[i]
} else if (E[i]<C4[i]){
W4[i]=1
A4[i]=(C4[i]-E[i])/(1-E[i]*C4[i])
S4[i]=A4[i]
} else if (G4[i]<E[i]){
W4[i]=0
B4[i]=(E[i]-G4[i])/(E[i]*G4[i]-1)
S4[i]=B4[i]
} else if (1<E[i] & E[i]<G4[i]){
W4[i]=0
B4[i]=0
S4[i]=B4[i]
}
}
}
# Single term candidates propose ends
# Two term parties propose begins
N5[1]=1
for (i in 2:k){
j=2*i
if (X[j-1]<X[j]){
C5[i]=X[j-1]/X[j]
G5[i]=1/C5[i]
```

```
} else if (X[j-1]>X[j]){
C5[i]=X[j]/X[j-1]
G5[i]=1/C5[i]
}
if (N5[i-1]==1){
if (W5[i-1]==1){
if (E[i]*X[j-3]>X[j]){
W5[i]=1
A5[i]=A5[i-1]
S5[i]=A5[i]
V5[i]=X[j-3]
N5[i]=N5[i-1]+1
} else if (E[i]*X[j-3]<X[j]){
W5[i]=0
B5[i]=1
S5[i]=B5[i]
V5[i]=X[j]
N5[i]=1
}
} else if (W5[i-1]==0){
if (E[i]*X[j-2]>X[j-1]){
W5[i]=0
B5[i]=B5[i-1]
S5[i]=B5[i]
V5[i]=X[j-2]
N5[i]=N5[i-1]+1
} else if
(E[i]*X[j-2]<X[j-1]){
W5[i]=1
A5[i]=1
S5[i]=A5[i]
V5[i]=X[j-1]
N5[i]=1
}
}
} else if (N5[i-1]==2){
if (W5[i-1]==1){
if (C5[i]<E[i] & E[i]<1){
W5[i]=0
B5[i]=1-(E[i]-C5[i])/(1-E[i]*C5[i])
S5[i]=B5[i]
if (X[j-1]<X[j]){
V5[i]=X[j]
} else {
V5[i]=X[j-1]
}
N5[i]=1
} else if (E[i]<C5[i]){
W5[i]=0
B5[i]=1
S5[i]=B5[i]
if (X[j-1]<X[j]){
V5[i]=X[j]
} else {
V5[i]=X[j-1]
}
N5[i]=1
} else if (G5[i]<E[i]){
W5[i]=1
A5[i]=1
S5[i]=A5[i]
if (X[j-1]<X[j]){
V5[i]=X[j]
} else {
V5[i]=X[j-1]
}
N5[i]=1
} else if (1<E[i] & E[i]<G5[i]){
W5[i]=1
A5[i]=1-(G5[i]-E[i])/(E[i]*G5[i]-1)
S5[i]=A5[i]
if (X[j-1]<X[j]){
V5[i]=X[j]
} else {
V5[i]=X[j-1]
}
N5[i]=1
}
} else if (W5[i-1]==0){
if (C5[i]<E[i] & E[i]<1){
W5[i]=1
```

```
A5[i]=1-(E[i]-C5[i])/(1-E[i]*C5[i])
S5[i]=A5[i]
if (X[j-1]<X[j]){
V5[i]=X[j]
} else {
V5[i]=X[j-1]
}
N5[i]=1
} else if (E[i]<C5[i]){
W5[i]=1
A5[i]=1
S5[i]=A5[i]
if (X[j-1]<X[j]){
V5[i]=X[j]
} else {
V5[i]=X[j-1]
}
N5[i]=1
} else if (G5[i]<E[i]){
W5[i]=0
B5[i]=1
S5[i]=B5[i]
if (X[j-1]<X[j]){
V5[i]=X[j]
} else {
V5[i]=X[j-1]
}
N5[i]=1
} else if (1<E[i] & E[i]<G5[i]){
W5[i]=0
B5[i]=1-(G5[i]-E[i])/(E[i]*G5[i]-1)
S5[i]=B5[i]
if (X[j-1]<X[j]){
V5[i]=X[j]
} else {
V5[i]=X[j-1]
}
N5[i]=1
}
}
}
}
# Two term parties propose ends
# Two term candidates propose begins
N6[1]=1
for (i in 2:k){
j=2*i
if (X[j-1]<X[j]){
C6[i]=X[j-1]/X[j]
G6[i]=1/C6[i]
} else if (X[j-1]>X[j]){
C6[i]=X[j]/X[j-1]
G6[i]=1/C6[i]
}
if (N6[i-1]==1){
if (W6[i-1]==1){
if (E[i]*X[j-3]>X[j]){
W6[i]=1
A6[i]=A6[i-1]
S6[i]=A6[i]
V6[i]=X[j-3]
N6[i]=N6[i-1]+1
} else if (E[i]*X[j-3]<X[j]){
W6[i]=0
u=runif(1,0,1)
if (u<0.5){
B6[i]=0
} else {
r=E[i]*X[j-3]
s=X[j]
v=runif(1,min=r,max=s)
B6[i]=(v-E[i]*X[j-3])/(X[j]-E[i]*X[j-3])
}
S6[i]=B6[i]
V6[i]=X[j]
N6[i]=1
}
} else if (W6[i-1]==0){
if (E[i]*X[j-2]>X[j-1]){
W6[i]=0
B6[i]=B6[i-1]
```

```
S6[i]=B6[i]
V6[i]=X[j-2]
N6[i]=N6[i-1]+1
} else if
(E[i]*X[j-2]<X[j-1]){
W6[i]=1
u=runif(1,0,1)
if (u<0.5){
A6[i]=0
} else {
r=E[i]*X[j-2]
s=X[j-1]
v=runif(1,min=r,max=s)
A6[i]=(v-E[i]*X[j-2])/(X[j-1]-E[i]*X[j-2])
}
S6[i]=A6[i]
V6[i]=X[j-1]
N6[i]=1
}
}
} else if (N6[i-1]==2){
if (W6[i-1]==1){
if (C6[i]<E[i] & E[i]<1){
W6[i]=0
B6[i]=0
S6[i]=B6[i]
if (X[j-1]<X[j]){
V6[i]=X[j]
} else {
V6[i]=X[j-1]
}
N6[i]=1
} else if (E[i]<C6[i]){
W6[i]=0
B6[i]=(C6[i]-E[i])/(1-E[i]*C6[i])
S6[i]=B6[i]
if (X[j-1]<X[j]){
V6[i]=X[j]
} else {
V6[i]=X[j-1]
}
N6[i]=1
} else if (G6[i]<E[i]){
W6[i]=1
A6[i]=(E[i]-G6[i])/(E[i]*G6[i]-1)
S6[i]=A6[i]
if (X[j-1]<X[j]){
V6[i]=X[j]
} else {
V6[i]=X[j-1]
}
N6[i]=1
} else if (1<E[i] & E[i]<G6[i]){
W6[i]=1
A6[i]=0
S6[i]=A6[i]
if (X[j-1]<X[j]){
V6[i]=X[j]
} else {
V6[i]=X[j-1]
}
N6[i]=1
}
} else if (W6[i-1]==0){
if (C6[i]<E[i] & E[i]<1){
W6[i]=1
A6[i]=0
S6[i]=A6[i]
if (X[j-1]<X[j]){
V6[i]=X[j]
} else {
V6[i]=X[j-1]
}
N6[i]=1
} else if (E[i]<C6[i]){
W6[i]=1
A6[i]=(C6[i]-E[i])/(1-E[i]*C6[i])
S6[i]=A6[i]
if (X[j-1]<X[j]){
V6[i]=X[j]
} else {
```

```
V6[i]=X[j-1]
}
N6[i]=1
} else if (G6[i]<E[i]){
W6[i]=0
B6[i]=(E[i]-G6[i])/(E[i]*G6[i]-1)
S6[i]=B6[i]
if (X[j-1]<X[j]){
V6[i]=X[j]
} else {
V6[i]=X[j-1]
}
N6[i]=1
} else if (1<E[i] & E[i]<G6[i]){
W6[i]=0
B6[i]=0
S6[i]=B6[i]
if (X[j-1]<X[j]){
V6[i]=X[j]
} else {
V6[i]=X[j-1]
}
N6[i]=1
}
}
}
}
# Two term candidates propose ends
# Stationary distribution plots begin
P1[1]=W1[1]
P2[1]=W2[1]
P3[1]=W3[1]
P4[1]=W4[1]
P5[1]=W5[1]
P6[1]=W6[1]
for (i in 2:k)
{
P1[i]=P1[i-1]+W1[i]
P2[i]=P2[i-1]+W2[i]
P3[i]=P3[i-1]+W3[i]
P4[i]=P4[i-1]+W4[i]
P5[i]=P5[i-1]+W5[i]
P6[i]=P6[i-1]+W6[i]
}
# Average number of victories per period
for (i in 1:k)
{
K1[i]=P1[i]/i
K2[i]=P2[i]/i
K3[i]=P3[i]/i
K4[i]=P4[i]/i
K5[i]=P5[i]/i
K6[i]=P6[i]/i
}
plot(K2, type="l", lty=1, lwd=2, col="red", xlab="Periods", ylab="Average no. of victories of A", xlim=c(0, 50), ylim=c(0, 0.7))
points(K3, type="l", lty=1, lwd=2, col="orange1")
points(K4, type="l", lty=2, lwd=2, col="chocolate4")
points(K5, type="l", lty=1, lwd=2, col="yellowgreen")
points(K6, type="l", lty=2, lwd=2, col="mediumblue")
points(K1, type="l", lty=1, lwd=2, col="black")
legend("topright", legend=c("Single term", "Single term PP", "Single term CP", "Two terms", "Two terms PP", "Two terms CP"), fill=c("black", "orange1", "chocolate4", "red", "yellowgreen", "mediumblue"), bty="n")
# Stationary distribution plots end
```